On the Fading Paper Achievable Region of the Fading MIMO Broadcast Channel

Amir Bennatan, *Member, IEEE* and David Burshtein, *Senior Member, IEEE*

Abstract

We consider transmission over the ergodic fading multi-antenna broadcast (MIMO-BC) channel with partial channel state information at the transmitter and full information at the receiver. Over the equivalent *non*-fading channel, capacity has recently been shown to be achievable using transmission schemes that were designed for the "dirty paper" channel. We focus on a similar "fading paper" model. The evaluation of the fading paper capacity is difficult to obtain. We confine ourselves to the *linear-assignment* capacity, which we define, and use convex analysis methods to prove that its maximizing distribution is Gaussian. We compare our fading-paper transmission to an application of dirty paper coding that ignores the partial state information and assumes the channel is fixed at the average fade. We show that a gain is easily achieved by appropriately exploiting the information. We also consider a cooperative upper bound on the sum-rate capacity as suggested by Sato. We present a numeric example that indicates that our scheme is capable of realizing much of this upper bound.

Index Terms

Broadcast channel, Dirty paper, MIMO, Sato bound

I. Introduction

The multiple-antenna Gaussian broadcast channel has recently been the subject of intense research. This surge of interest was spurred by the seminal work of Caire and Shamai [6], who suggested an achievable region for this channel based on dirty-paper coding. Recently, this region was shown by Weingarten $et\ al.$ [30] to exhaust the capacity region of the channel.

However, the channel model examined in [6] assumes that the fading coefficients of the MIMO channel are fixed and known to both the transmitter and the receiver. In several realistic settings, the coefficients fluctuate over time. They are estimated at the receiver and are fed back to the transmitter. At best, we can assume that the transmitter has a rough, outdated estimate of the coefficients.

Manuscript submitted to IEEE Transactions on Information Theory: June 2006, revised: July 2007, to be published. This research was supported by the Israel Science Foundation, grant no. 927/05. The material in this paper was presented at the 44th Annual Allerton Conference on Communications, Control and Computing, Monticello, IL, September 2006.

A. Bennatan was with the School of Electrical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel. He is now with the Program in Applied and Computational Mathematics (PACM) at Princeton University, Princeton, NJ 08544 USA (e-mail: abn@math.princeton.edu).

D. Burshtein is with the School of Electrical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel (e-mail: burstyn@eng.tau.ac.il).

Telatar $et\ al.$ [27], in his work on the single-user MIMO channel, focused on a setting where the transmitter has zero knowledge of the fading coefficients. In a broadcast setting, this problem is typically uninteresting because its solution is often trivial. In Appendix I, we will see such a setting where time-sharing (TDMA) is the best that can be achieved. However, in a realistic setting, the transmitter has *some* knowledge of the channel to each of the users. This knowledge can be modelled as channel distribution information[1].

We assume an ergodic channel, in the sense that a new channel realization is obtained at each time instance. However, the channel distribution, which is known to the transmitter, remains fixed for the duration of the transmission.

The analysis of ergodic broadcast channels was initiated by Cover [10]. The capacity of such channels is known only in special cases, where the signals to the users can be ordered according to their "strength". A large class of such channels, known as "more capable" channels, was considered by El Gamal [12], who also evaluated the capacity in this case. This class contains "degraded" and "less noisy" channels as special cases [12].

Tuninetti and Shamai [28] considered the fading *scalar* broadcast channel, which is a special case of the fading MIMO-BC channel obtained by setting the number of antennas at the transmitter and receivers to one. They showed that this channel is not "more capable" in general. They nonetheless evaluated the "more capable" region as defined by [12]. This region is still achievable despite the channel being not "more capable", although it is only an inner bound and does not exhaust the entire capacity region.

Jafar $et\ al.$ [16] considered the fading MISO-BC, characterized by receivers that have only one antenna each. They considered the case when the distribution of the fading coefficients is isotropic. In this case, they proved that the capacity region collapses to that of the above fading scalar channel. Lapidoth [21] examined a similar two-user fading MISO-BC channel, and demonstrated that at the limit of high SNR, a significant loss is incurred as a result of the unavailability of precise channel state information at the transmitter. Sharif and Hassibi [25] proposed a beamforming transmission approach for the case when the knowledge available to the transmitter is the collection of SINR values available to each of the receivers.

The fading MIMO-BC channel, being not "more capable" in general, is difficult to analyze. In this paper we focus on an achievable region which is modelled on the dirty paper region of Caire and Shamai [6]. Our development uses a *fading-paper* approach which is a generalization of the dirty-paper approach of [6]. A fading paper solution was previously considered for a wideband fading channel in [3], although they assumed an interference which is known only causally, unlike the dirty paper problem of Costa. The proof of [30] does not apply to the fading MIMO-BC capacity region, so that the fading paper approach is not guaranteed to be optimal. Furthermore, the capacity of the fading-paper channel is in general not known. We focus on its *linear-assignment* capacity, which we define. We use convex-optimization methods to prove that a Gaussian

[1]A different model was proposed by Jindal [18] and Caire [5], who incorporated the feedback from the receiver into the channel model.

distribution achieves this capacity.

We compare the rate region achieved by this approach to the region that is achievable by a dirty-paper scheme that ignores the available channel state information and assumes that the channel is fixed at its average. We show that a substantial benefit is easily achieved by appropriately exploiting the available information.

This paper is organized as follows. We begin with some background in Sec. II. We define our notation and the channel model, discuss the dirty-paper channel and its application to transmission over the non-fading MIMO-BC channel. In Sec. III we discuss the fading-paper generalization of the dirty-paper channel, define the linear-assignment capacity and discuss its maximizing distribution. In Sec. IV we define a region that is achievable using linear-assignment fading-paper transmission methods. We also compare this region to that of dirty-paper based transmission that assumes the channel is fixed at its average. In Sec. V we present ideas for further research and conclude the paper.

II. Background

A. Notation

E_H denotes the expectation over the random variable H. Matrices are denoted by upper-case letters, with bold indicating realizations of random variables (e.g. $\mathbf{H}$ is the realization of H). Vector values are denoted in boldface and scalar values are denoted in normal typeface. With both, lower-case letters denote the realizations of random variables ($\mathbf{y}$ is a realization of $\mathbf{Y}$ and y is a realization of Y).

The inner product of two equal-dimension matrices $A, B \in \mathbb{R}^{M \times N}$ is defined by,

$$< A, B > \triangleq \sum_{m=1}^{M} \sum_{n=1}^{N} A_{m,n} B_{m,n} = \mathrm{tr}[A \cdot B^T]$$

$\mathbb{R}_+$ denotes the non-negative real numbers and $\mathbb{R}_{++}$ the positive real numbers.

B. System Model

We consider a broadcast channel with L users. The transmitter has M transmit antennas and user l has N_l antennas. For simplicity we assume that all signals are real-valued.

The channel output $\mathbf{Y}_t^{(l)}$ observed by receiver u at a discrete time instance t is given by,

$$\mathbf{Y}_t^{(l)} = H_t^{(l)} \cdot \mathbf{X}_t + \mathbf{Z}_t^{(l)}$$

$\mathbf{Y}_t^{(l)}$ is a $N_l \times 1$ column vector. $\mathbf{H}_t^{(l)}$ is a random $N_l \times M$ matrix denoting the channel transition matrix. We assume that instances of $H_t^{(l)}$ are independent over time (for different values of t) and between users (i.e., for different values of l). As noted in Sec. I, we assume that this matrix is known to the receiver, and in our subsequent analysis, we consider it as part of the channel output. $\mathbf{X}_t$ is an $M \times 1$ column vector denoting the transmitted signal. $\mathbf{Z}_t^{(l)}$ denotes Gaussian noise, distributed as a N_l-dimensional zero-mean Gaussian random variable with identity covariance matrix $\mathbf{I}$ [2].

[2]If the noise's covariance matrix is not $\mathbf{I}$, we can multiply $\mathbf{Y}_t^{(l)}$ by the inverse of the square root of of the matrix and obtain an equivalent channel that does agree with this model.

In the sequel, for simplicity, we will drop the time index t. We assume that the transmitter is subject to an average power constraint P. That is, we require,

$$\mathrm{E}\,\mathrm{tr}(\mathbf{X}\mathbf{X}^T) \leq P$$

The only assumption we make on the distribution of $H^{(l)}$ is that is has finite energy, i.e. $\mathrm{E} < H^{(l)}, H^{(l)} >$ is finite.

C. Dirty Paper Channels

The dirty-paper channel was first considered by Costa [8]. It is defined by

$$Y = X + S + Z \tag{1}$$

The channel input X is subject to a power constraint P, i.e. The noise Z is distributed as a zero-mean Gaussian variable with variance $\sigma_Z^2 > 0$. S is interference, known to the transmitter but not to the receiver.

Costa obtained the remarkable result that the interference, despite being known only to the encoder, incurs no loss of capacity in comparison with the standard interference-free channel. Costa assumed that S is Gaussian i.i.d distributed. This result was extended in [7] and [13] to arbitrarily distributed interference. Costa's result was further extended to the Gaussian MIMO channel by Yu $et\ al.$ [31]. With this channel model, vector $\mathbf{Y}$, $\mathbf{S}$, $\mathbf{X}$ and $\mathbf{Z}$ replace the above scalar equivalents, $\mathbf{Z}$ being a zero-mean Gaussian random vector with nonsingular covariance matrix Σ_Z [3].

In Sec. II-D we will consider dirty-paper in the context of transmission over nonfading MIMO-BC channels. In that context, it will be useful to consider the following variation of (1) (using vector substitutes for Y, S, X and Z),

$$\mathbf{Y} = \mathbf{H}(\mathbf{X} + \mathbf{S}) + \mathbf{Z} \tag{2}$$

where $\mathbf{S}$ and $\mathbf{X}$ are M dimensional, $\mathbf{Y}$ and $\mathbf{Z}$ are N dimensional, and $\mathbf{H}$ is an $N \times M$ fixed channel matrix[4]. We assume this formulation of the dirty-paper problem throughout the rest of this paper. Once again, the capacity coincides with that of the corresponding no-interference channel, whose output $\hat{\mathbf{Y}}$ is given by,

$$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{X} + \mathbf{Z} \tag{3}$$

The dirty-paper channel is an instance of the more general class of *side-information* channels, first considered by Shannon [24]. Such channels are characterized by an input X, output Y and state-dependent transition probabilities $\Pr[y|x, s]$ where the channel state S is i.i.d., known to the transmitter and unknown to the receiver. In the context of (1), the interference S constitutes the channel state.

[3]Note that unlike the fading MIMO-BC model of Sec. III-A, we find it more convenient to allow $\Sigma_Z \neq \mathbf{I}$ in this context of the vector dirty-paper channel.

[4]The matrix $\mathbf{H}$ is denoted in bold since in the next section it will be a realization of a random variable.

Shannon [24] considered the case of the state sequence being known only causally. Kusnetsov and Tsybakov [20] were the first to consider the case of state sequence known non-causally, and Gel'fand and Pinsker [14] obtained the capacity formula for this case. The capacity of this channel is given by

$$C = \sup_{\Pr[u \mid s], \mathrm{f}(\cdot)} \{I(U;Y) - I(U;S)\} \tag{4}$$

where U is an auxiliary random variable with conditional distribution $\Pr[u \mid s]$ and $\mathrm{f}(\cdot)$ is a deterministic function, such that the transmitted signal X is given by $X = \mathrm{f}(S, U)$.

In [31], the capacity of the dirty-paper channel was obtained from (4) using an auxiliary random variable $\mathbf{U}$ given by $\mathbf{U} = \mathbf{F}\cdot\mathbf{S}+\mathbf{X}$, where $\mathbf{F}$ is a fixed matrix[5] and $\mathbf{X}$ is a zero-mean Gaussian-distributed random-variable, independent of $\mathbf{S}$. The use of $\mathbf{X}$ has a dual role. First, it is a component in the definition of the transition probabilities $\Pr[\mathbf{u} \mid \mathbf{s}]$. Second, given $\mathbf{U}$ and $\mathbf{S}$, the transmitted signal satisfies $\mathbf{f}(\mathbf{U}, \mathbf{S}) \triangleq \mathbf{U} - \mathbf{F} \cdot \mathbf{S} = \mathbf{X}$. The covariance matrix Σ_X of $\mathbf{X}$ is determined as in the no-interfence channel (see e.g. [9]). An expression for $\mathbf{F}$ was developed by Yu and Cioffi [32]. In this paper, we use the following, equivalent expression:

$$\mathbf{F} = \Sigma_X \mathbf{H}^T (\mathbf{H} \Sigma_X \mathbf{H}^T + \Sigma_Z)^{-1} \mathbf{H} \tag{5}$$

A proof that this choice of $\mathbf{F}$ indeed achieves the no-inteference capacity is provided in Appendix II. This proof is different from the proof of [32], and is provided primarily for completeness.

Costa [8] and Yu [31] obtained their results using random codes and maximum-likelihood decoding. Zamir $et\ al.$ [33] and Bennatan $et\ al.$ [1] have presented practical methods for transmitting at rates that approach the above computed capacities. Their approaches were developed for the scalar dirty-paper channel, but can easily be adapted to the MIMO setting [1][Sec. VII].

D. The Dirty-Paper Achievable Region

In their construction for the non-fading MIMO broadcast channel, Caire and Shamai [6] used dirty-paper coding to transmit in the following way. The transmitted signal $\mathbf{X}$ is constructed as the vector sum of L signals $\mathbf{X}_1, ..., \mathbf{X}_L$, where $\mathbf{X}_l$ contains the transmitted signal to user l. Each user is also allotted a virtual power constraint P_l such that $\sum_{l=1}^{L} P_l = P$. Using dirty-paper coding, the transmitter can generate the signal $\mathbf{X}_l$ such that the interference generated by $\mathbf{X}_1, ..., \mathbf{X}_{l-1}$ is effectively pre-subtracted. More precisely, encoding proceeds in the following way,

1) The transmitter begins by selecting a codeword $\mathbf{c}_1$ for user 1.
2) It then proceeds to determine the signal for user 2. It constructs the signal $\mathbf{X}_2$ for user 2 using a dirty-paper transmission scheme, making use of its full non-causal knowledge of $\mathbf{c}_1$ and treating it as known interference (in lieu of S in (1)).
3) The signals $\mathbf{X}_3, ..., \mathbf{X}_L$ are constructed in a similar manner. When constructing the signal to user l, the signal $\mathbf{S}^{(l)} \triangleq \mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_{l-1}$ is treated as non-causally known interference.

[5]We denote the matrix $\mathbf{F}$ in bold throughout the paper in order to distinguish it from the functional $F(q, Q)$.

The operation of the receivers mirrors the above transmission scheme. Receiver l applies dirty-paper decoding, effectively cancelling the interference generated by $\mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_{l-1}$ but treating $\mathbf{X}_{l+1} + ... + \mathbf{X}_L$ as part of the unknown noise (alongside $\mathbf{Z}$).

The above transmission strategy defines an achievable rate region for the Gaussian MIMO broadcast channel. This region is a function of the virtual power constraints P_l imposed on the users. Furthermore, it is a function of the covariance matrices $\Sigma_X^{(l)}$ by which the various codebooks for the signals $\mathbf{X}_l$ are randomly generated. It is also a function of the ordering of the users. The convex-hull of the union of all regions obtained in this way constitutes the dirty-paper achievable region $\mathcal{C}_{DPC}(P)$. In [30], this region was shown to exhaust the MIMO broadcast capacity region.

However, the application of dirty-paper transmission methods in the above algorithm is heavily reliant on the availability of precise knowledge of the fixed channel matrices $\{\mathbf{H}^{(l)}\}_{l=1}^L$ at the transmitter. Without these, the pre-subtraction of the signals $\{\mathbf{X}_i\}_{i<l}$, when constructing $\mathbf{X}_l$, is not possible.

III. The Fading-Paper Problem

A. Channel Model

The fading-paper channel is an adaptation of the dirty-paper model (as expressed in (2)) of Sec. II-C, designed to account for the absence of channel state information at the receiver. The channel is defined by,

$$\mathbf{Y} = H(\mathbf{X} + \mathbf{S}) + \mathbf{Z} \tag{6}$$

Unlike the case in (2), the channel matrix is random and is know to the receiver but not to the transmitter. The pair $(\mathbf{Y}, H)$ constitutes the channel output, where $\mathbf{Y}$ is the channel observation and H is the channel matrix.

The channel transition probabilities are also a function of the distribution of the interference $\mathbf{S}$ and of the channel matrix H. In this paper, we assume $\mathbf{S}$ to be a zero-mean Gaussian distributed random variable with covariance Σ_S. As noted in Sec. II-B, we make no assumptions on the distribution of H, beyond it having finite energy. Following the discussion of side-information channels in Sec. II-C, the capacity of the fading-paper channel is given by,

$$C = \sup_{\Pr[\mathbf{u}\,|\,\mathbf{s}],\mathbf{f}(\cdot)} \{I(\mathbf{U}; \mathbf{Y}, H) - I(\mathbf{U}; \mathbf{S})\} \tag{7}$$

where $\mathbf{U}$ is an auxiliary random variable whose joint distribution with $\mathbf{S}$ can be obtained via $\Pr[\mathbf{u} \mid \mathbf{s}]$. $\mathbf{f}(\cdot)$ is a vector-valued deterministic function, such that the transmitted signal $\mathbf{X}$ is given by $\mathbf{X} = \mathbf{f}(\mathbf{U}, \mathbf{S})$.

Note that for any particular choice of $\Pr[\mathbf{u} \mid \mathbf{s}]$ and $\mathbf{f}(\cdot)$, the contents of the braces are an achievable transmission rate over the channel,

$$R_{\text{achievable}} = I(\mathbf{U}; \mathbf{Y}, H) - I(\mathbf{U}; \mathbf{S}) \tag{8}$$

B. The Linear-Assignment Capacity

In this paper, we focus on a subset of achievable rates for the fading-paper channels, modelled on the dirty-paper capacity-achieving assignment for $\mathbf{U}$ and $\mathbf{f}(\cdot)$. That is, we focus on an auxiliary random variable $\mathbf{U}$ given by

$$\mathbf{U} = \mathbf{F} \cdot \mathbf{S} + \mathbf{X} \tag{9}$$

where $\mathbf{F}$ is some arbitrary real-valued $M \times M$ matrix, and $\mathbf{X}$ is an arbitrary zero-mean random-variable, which may depend on $\mathbf{S}$. We define $\mathbf{f}(\mathbf{u}, \mathbf{s}) = \mathbf{u} - \mathbf{F}\mathbf{s}$. We refer to such an assignment as a *linear assignment*. We call the maximum in (7), when restricted to such assignments, the *linear assignment capacity*.

Linear assignments may equivalently be defined as follows. A linear assignment is characterized by an arbitrary zero-mean M-dimensional random variable $\mathbf{U}$ (recall that M is the dimension of $\mathbf{X}$ and $\mathbf{S}$), which may be dependent on $\mathbf{S}$, and an arbitrary real-valued $M \times M$ matrix $\mathbf{F}$. In the context of (4), $\mathbf{U}$ corresponds to the auxiliary variable $\mathbf{U}$ and $\mathbf{f}(\cdot,\cdot)$ is defined by $\mathbf{f}(\mathbf{u}, \mathbf{s}) = \mathbf{u} - \mathbf{F}\mathbf{s}$. A set $\mathbf{U}, \mathbf{F}$ and $\mathbf{f}(\cdot,\cdot)$ given by the first definition straightforwardly satisfies the conditions of the second definition. To see that the reverse holds, observe that we have allowed $\mathbf{X}$ to be completely arbitrary. In particular, we have in no way required $\mathbf{X}$ to be Gaussian or independent of $\mathbf{S}$. Thus, given a pair $\mathbf{U}$ and $\mathbf{F}$ corresponding to the second definition, we may define $\mathbf{X} = \mathbf{U} - \mathbf{F} \cdot \mathbf{S}$ and the resulting set $\mathbf{U}, \mathbf{X}, \mathbf{F}$ and $\mathbf{f}(\cdot,\cdot)$ coincides with the first definition.

The optimality of linear assignments for the dirty-paper problem of Sec. II-C is obtained from the fact that their maximum achievable rate coincides with the capacity of the corresponding no-interference channel. This is clearly the best we can hope for, and thus such assignments achieve capacity. With fading-paper, the achievable rate with linear assignments is in general strictly *below* the no-interference upper-bound. Thus, it is not known whether it is optimal.

In our above definition of linear assignments, we left the distribution of $\mathbf{X}$ undefined. Specifically (as noted above), we did not insist on $\mathbf{X}$ to be Gaussian, and did not insist on it being independent of $\mathbf{S}$, as we did in Sec. II-C when we discussed the capacity-achieving assignment for the dirty-paper channel. However, the following theorem establishes the optimality of a Gaussian-distributed $\mathbf{X}$. In Sec. IV we will show that we may also assume $\mathbf{X}$ to be independent of $\mathbf{S}$.

In the following theorem, we assume the following regularity conditions:

1) We assume that the expectations (29), (30), (31) and (32) (defined below), exist and are finite. Note that this condition is satisfied, for example, when the distribution of H is discrete and takes a finite set of values.
2) We assume that the covariance matrix of the vector $(\mathbf{S}, \mathbf{U})$,

$$\begin{pmatrix} \Sigma_S & \Sigma_{S,U} \\ \Sigma_{S,U}^T & \Sigma_U \end{pmatrix} \tag{10}$$

is nonsingular (i.e., it is a positive definite matrix). Note that this also implies that Σ_S is nonsingular, being a principal submatrix of $\mathrm{Cov}(\mathbf{U},\mathbf{S})$. Since,

$$\begin{pmatrix} \mathbf{S} \\ \mathbf{U} \end{pmatrix} = \begin{pmatrix} I & 0 \\ F & I \end{pmatrix} \begin{pmatrix} \mathbf{S} \\ \mathbf{X} \end{pmatrix}$$

and since the matrix on the right hand side of the last equation is nonsingular, a sufficient condition that (10) is nonsingular is that $\det(\Sigma_S) > 0$, $\det(\Sigma_X) > 0$ and $\det(\Sigma_X - \Sigma_{S,X}^T \Sigma_S^{-1} \Sigma_{S,X}) > 0$ (Σ_X and $\Sigma_{S,X}$ are the covariance of $\mathbf{X}$ and the cross-covariance of $\mathbf{S}$ and $\mathbf{X}$, respectively).

3) We assume an arbitrary density $q(\mathbf{u} \,|\, \mathbf{s})$ with respect to the Lebesgue measure.

Definition 1: Given a linear assignment, the collection of matrices $\Sigma_S, \Sigma_{S,X}, \Sigma_X$ and $\mathbf{F}$ is called its *setting*.

Theorem 1: Assume the above-mentioned regularity conditions. For any fixed setting, the linear-assignment capacity (as defined above) is achieved by a choice of $\mathbf{X}$ that is jointly Gaussian with $\mathbf{S}$.

Proof: We begin with a brief outline of the proof. We consider (8) as a function of the density $q(\mathbf{u}|\mathbf{s})$ and of $Q(\mathbf{u} \,|\, \mathbf{y}, \mathbf{H})$, defined below. We then seek to show that $q_G(\cdot)$ and $Q_G(\cdot)$, corresponding to a joint-Gaussian choice of $\mathbf{X}$ and $\mathbf{S}$, maximize (8). To do so, we pose the problem as a concave constrained maximization problem, and show that q_G and Q_G admit Lagrange multipliers.

We now rewrite (8) as $F(q, Q)$, given by[6],

$$\begin{aligned} F(q,Q) \quad &\triangleq \quad \int_{\mathbf{s}\in\mathbb{R}^M} \int_{\mathbf{u}\in\mathbb{R}^M} \int_{\mathbf{y}\in\mathbb{R}^N} \int_{\mathbf{H}\in\mathcal{R}_H} f_{\mathbf{S}}(\mathbf{s}) f_{Y,H\,|\,S,X}(\mathbf{y},\mathbf{H} \,|\, \mathbf{s}, \mathbf{x} = \mathbf{f}(\mathbf{u},\mathbf{s})) q(\mathbf{u} \,|\, \mathbf{s}) \cdot \\ & \quad \cdot \log \frac{Q(\mathbf{u} \,|\, \mathbf{y}, \mathbf{H})}{q(\mathbf{u} \,|\, \mathbf{s})} \, d\mathbf{H} \, d\mathbf{y} \, d\mathbf{u} \, d\mathbf{s} \end{aligned} \tag{11}$$

Recall that M and N are the dimensions of $\mathbf{S}$ and $\mathbf{Y}$, respectively. We also denote by $\mathcal{R}_H$ the support region of the random variable $\mathbf{H}$. $Q(\mathbf{u}|\mathbf{y},\mathbf{H})$ is the conditional distribution of the above-defined $\mathbf{U}$ given the channel output $\mathbf{Y}$ and the signal fade $\mathbf{H}$. $f_S(\mathbf{s})$ is the density of $\mathbf{S}$ and $f_{Y,H\,|\,S,X}(\mathbf{y},\mathbf{H}|\mathbf{s},\mathbf{x})$ is the conditional density of $\mathbf{Y}$ and $\mathbf{H}$ given the transmitted $\mathbf{x}$ and interference $\mathbf{s}$.

Since we make no assumptions on the distribution of $\mathbf{H}$, the existence of this density is not guaranteed. However, the generalization to the case when the density does not exist is straightforward. In the sequel, we drop the subscripts and denote the densities by $f(\mathbf{s})$ and $f(\mathbf{y},\mathbf{H}|\mathbf{s},\mathbf{x})$. Note that $f(\mathbf{s})$ should not be confused with the previously defined $\mathbf{f}(\mathbf{u},\mathbf{s})$.

We defined $Q(\mathbf{u} \mid \mathbf{y}, \mathbf{H})$ in (11) to be the conditional density of the above-defined $\mathbf{U}$ given the channel output $\mathbf{Y}$ and the signal fade $\mathbf{H}$. Actually, in the sequel we find it convenient to relax this requirement and consider $F(q, Q)$ for arbitrary probability densities $Q(\mathbf{u} \mid \mathbf{y}, \mathbf{H})$. However, the pair q and Q that *maximizes* $F(q, Q)$ *will* satisfy the requirement. In this we follow the example of [17].

For given Σ_S, Σ_X and $\Sigma_{S,X}$, let $q_G(\mathbf{u}\,|\,\mathbf{s})$ and $Q_G(\mathbf{u}\,|\,\mathbf{y},\mathbf{H})$ denote the conditional densities corresponding to the choice of $\mathbf{X}$ that is jointly-Gaussian with $\mathbf{S}$. Our objective is to show that q_G and Q_G maximize $F(q, Q)$.

[6] This definition is an adaptation of a similar definition by Heegard and El Gamal [17]

$F(q, Q)$ as defined in (11) is jointly-concave in its arguments. Thus we may wish to apply methods from the theory of convex optimization to maximize it. Formally, we seek to solve the following constrained problem

$$\max_{q,Q} F(q, Q) \text{ subject to} \tag{12}$$

$$\int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M} f(\mathbf{s})q(\mathbf{u}\,|\,\mathbf{s})\left[\mathbf{f}(\mathbf{u},\mathbf{s})\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T\right]\,d\mathbf{u}\,d\mathbf{s} = \Sigma_X \tag{13}$$

$$\int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M} f(\mathbf{s})q(\mathbf{u}\,|\,\mathbf{s})\left[\mathbf{s}\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T\right]\,d\mathbf{u}\,d\mathbf{s} = \Sigma_{S,X} \tag{14}$$

$$\int_{\mathbf{u}\in\mathbb{R}^M} q(\mathbf{u}\,|\,\mathbf{s})\,d\mathbf{u} = 1 \qquad \forall\mathbf{s}\in\mathbb{R}^M \tag{15}$$

$$\int_{\mathbf{u}\in\mathbb{R}^M} Q(\mathbf{u}\,|\,\mathbf{y},\mathbf{H})\,d\mathbf{u} = 1 \qquad \forall\mathbf{y}\in\mathbb{R}^N,\forall\mathbf{H}\in\mathcal{R}_H \tag{16}$$

Recall that Theorem 1 assumes a fixed setting. Thus, the matrices $\Sigma_S, \Sigma_{S,X}, \Sigma_X$ and $\mathbf{F}$ are assumed to be given and fixed. The maximization is performed over the set of distributions corresponding to these matrices, and our objective is to show that a Gaussian distribution is optimal. Optimization of the matrices themselves is beyond the scope of this proof (such optimization will be discussed in Sec. IV-B).

(13) and (14) are derived from the conditions Σ_X and $\Sigma_{S,X}$ on the transmitted signal $\mathbf{X}$. That is, recalling that $\mathbf{X} = \mathbf{f}(\mathbf{U}, \mathbf{S})$, they are equivalent to

$$\mathrm{E}\left[\mathbf{X}\cdot\mathbf{X}^T\right] = \Sigma_X, \quad \mathrm{E}\left[\mathbf{S}\cdot\mathbf{X}^T\right] = \Sigma_{S,X}$$

To further simplify our analysis, we allow the arguments q and Q of $F(q, Q)$ to be arbitrary nonnegative measurable functions. Constraints (15) and (16), compensate for this and ensure that the final result is a valid conditional distribution. Functions q and Q that satisfy constraints (13), (14), (15) and (16) are called *feasible*.

A straightforward approach to our optimization problem would appear to be to apply the Karush-Kuhn-Tucker (KKT) conditions to find the global maximum. In reality, this is slightly more involved because equations (15) and (16) involve an infinite number of constraints. Furthermore, the arguments of $F(q, Q)$ are functions rather than vectors. In [26], the necessity of the KKT conditions was proven under certain conditions. In this paper, we only require their sufficiency for convex functionals, which is easier to prove. Our proof is tailored to the setting of our particular problem. We begin by defining Lagrange multipliers.

Definition 2: Let q, Q be two positive-valued[7] feasible functions. Lagrange multipliers for q and Q are matrices $\Gamma, \Upsilon \in \mathbb{R}^{M\times M}$, and real-valued functions $\alpha(\mathbf{s}) : \mathbb{R}^M \to \mathbb{R}$ and $\beta(\mathbf{y}, \mathbf{H}) : \mathbb{R}^N \times \mathcal{R}_H \to \mathbb{R}$ such that,

$$\begin{gathered}\int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\,|\,\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))\left[\log\frac{Q(\mathbf{u}\,|\,\mathbf{y},\mathbf{H})}{q(\mathbf{u}\,|\,\mathbf{s})}-1\right]\,d\mathbf{H}\,d\mathbf{y}+\\ f(\mathbf{s}) < \Upsilon, \mathbf{f}(\mathbf{u},\mathbf{s})\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T > + f(\mathbf{s}) < \Gamma, \mathbf{s}\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T > + \alpha(\mathbf{s}) = 0\\ \forall\mathbf{s}\in\mathbb{R}^M, \forall\mathbf{u}\in\mathbb{R}^M\end{gathered} \tag{17}$$

$$\begin{gathered}\int_{\mathbf{s}\in\mathbb{R}^M} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\,|\,\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))\frac{q(\mathbf{u}\,|\,\mathbf{s})}{Q(\mathbf{u}\,|\,\mathbf{y},\mathbf{H})}\,d\mathbf{s} + \beta(\mathbf{y},\mathbf{H}) = 0\\ \forall\mathbf{u}\in\mathbb{R}^M, \forall\mathbf{y}\in\mathbb{R}^N, \forall\mathbf{H}\in\mathcal{R}_H\end{gathered} \tag{18}$$

[7]The condition that q and Q be positive-valued is required for the expressions that follow, which involve division by $Q(\mathbf{u}\,|\,\mathbf{y},\mathbf{H})$ and $q(\mathbf{u}\,|\,\mathbf{s})$, to be valid.

We say that two functions q and Q *admit* Lagrange multipliers if Lagrange multipliers that satisfy Definition 2 exist for them.

To obtain some motivation for (17) and (18), consider the formal Lagrangian, defined as

$$\mathcal{L}(q,Q;\,\Upsilon,\Gamma,\alpha,\beta) \triangleq F(q,Q) + <\Upsilon,\mathbf{E}(q)> + <\Gamma,\mathbf{C}(q)> + \int_{\mathbf{s}\in\mathbb{R}^M}\alpha(\mathbf{s})\cdot\int_{\mathbf{u}\in\mathbb{R}^M} q(\mathbf{u}\,|\,\mathbf{s})\,d\mathbf{u}\,d\mathbf{s} + \\ + \int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H}\beta(\mathbf{y},\mathbf{H})\cdot\int_{\mathbf{u}\in\mathbb{R}^M} Q(\mathbf{u}\,|\,\mathbf{y},\mathbf{H})\,d\mathbf{u}\,d\mathbf{H}\,d\mathbf{y} \tag{19}$$

where $\mathbf{E}(q)$ and $\mathbf{C}(q)$ are matrix-valued functionals given by the left-hand-side of (13) and (14). Formally differentiating $\mathcal{L}(q,Q;\,\Upsilon,\Gamma,\alpha,\beta)$ with respect to $q(\mathbf{u}\,|\,\mathbf{s})$ (for given $\mathbf{u}$ and $\mathbf{s}$) and comparing with zero, would render (17). Similarly, differentiating with respect to $Q(\mathbf{u}\,|\,\mathbf{y},\mathbf{H})$ (for given $\mathbf{u}$,$\mathbf{y}$ and $\mathbf{H}$), and comparing with zero, would render (18). However, the integrals in (19) are defined over unbounded sets, making their rigorous analysis difficult. We therefore prefer to avoid the use of (19), and rely on Definition 2 as the definition for Lagrange multipliers.

We are now ready for the following lemma,

Lemma 1: Let $q^\star$ and $Q^\star$ be a pair of positive-valued feasible functions for the problem (12). Assume once again that $Q^\star$ is the marginal distribution of $\mathbf{U}$ given $\mathbf{y}$ and $\mathbf{H}$, when the distribution of $\mathbf{U}$ is determined from the densities $f(\mathbf{s})$ and $q^\star(\mathbf{u}\,|\,\mathbf{s})$. If $q^\star$ and $Q^\star$ admit Lagrange multipliers, then they are a solution (i.e., achieve the global maximum) of (12).

A proof of Lemma 1 is provided in Appendix III. The proof is basically an application of well-known concepts from convex optimization theory. The proof of Theorem 1 now focuses on showing that the above defined q_G and Q_G admit Lagrange multipliers. We begin by providing the expressions for these two densities.

Recall once more that the setting of the problem (see Definition 1) is fixed. That is, we assume that Σ_S, $\Sigma_{S,X}$, Σ_X and $\mathbf{F}$ are given and fixed. Also recall that $\mathbf{U}$ is related to $\mathbf{S}$ and $\mathbf{X}$ through $\mathbf{U}=\mathbf{F}\mathbf{S}+\mathbf{X}$ and that q_G and Q_G correspond to a choice of $\mathbf{X}$ that is jointly-Gaussian with $\mathbf{S}$.

To obtain q_G, we observe that since $\mathbf{U}$ and $\mathbf{S}$ are jointly-Gaussian, the conditional distribution of $\mathbf{U}$ given $\mathbf{S}$ is also Gaussian, with mean $\mathbf{m}_{U\,|\,S}(\mathbf{s})$ and covariance $\Sigma_{U\,|\,S}$ given by (see e.g. [19]),

$$\begin{aligned} \mathbf{m}_{U\,|\,S}(\mathbf{s}) &= \mathrm{E}\mathbf{U} + \mathrm{Cov}(\mathbf{U},\mathbf{S})\cdot\Sigma_S^{-1}\cdot(\mathbf{s}-\mathrm{E}\mathbf{S}) \\ \Sigma_{U\,|\,S} &= \mathrm{Cov}(\mathbf{U}) - \mathrm{Cov}(\mathbf{U},\mathbf{S})\cdot\Sigma_S^{-1}\cdot\mathrm{Cov}(\mathbf{S},\mathbf{U}) \end{aligned}$$

Note that by our second regularity assumption (above), that the covariance of $(\mathbf{U},\mathbf{S})$ is nonsingular (positive definite), it follows that $\Sigma_{U\,|\,S}$ is also nonsingular[8].

Using $\mathbf{U}=\mathbf{F}\mathbf{S}+\mathbf{X}$ and $\mathrm{E}\mathbf{U}=\mathrm{E}\mathbf{S}=\mathbf{0}$, we obtain,

$$\mathbf{m}_{U\,|\,S}(\mathbf{s}) = J\mathbf{s}, \quad \text{where} \quad J \triangleq (\mathbf{F}\Sigma_S + \Sigma_{S,X}^T)\Sigma_S^{-1} \tag{20}$$

$$\Sigma_{U\,|\,S} = (\mathbf{F}\Sigma_S\mathbf{F}^T + \mathbf{F}\Sigma_{S,X} + \Sigma_{S,X}^T\mathbf{F}^T + \Sigma_X) - (\mathbf{F}\Sigma_S + \Sigma_{S,X}^T)\Sigma_S^{-1}(\mathbf{F}\Sigma_S + \Sigma_{S,X}^T)^T \tag{21}$$

[8] To see this, assume by contradiction that $\mathbf{v}\Sigma_{U\,|\,S}\mathbf{v}^T=0$ for some nonzero row vector $\mathbf{v}$. Thus, with probability 1 we would have $\mathbf{v}\cdot\mathbf{U}=\mathbf{v}\cdot\mathbf{m}_{U\,|\,S}(\mathbf{S})$, and therefore, using (20), $[\mathbf{v},-\mathbf{v}J]\cdot[\mathbf{U}^T,\mathbf{S}^T]^T=0$. This would imply that $\mathrm{Cov}(\mathbf{U},\mathbf{S})$ is singular.

Observe that J and $\Sigma_{U \mid S}$ are fixed matrix functions of the matrices Σ_X, Σ_S, $\Sigma_{S,X}$ and $\mathbf{F}$ that constitute the problem setting. Hence,

$$q_G(\mathbf{u} \mid \mathbf{s}) = \frac{1}{\sqrt{\det(2\pi\Sigma_{U \mid S})}} \exp(-\frac{1}{2}(\mathbf{u} - J\mathbf{s})^T \Sigma_{U \mid S}^{-1}(\mathbf{u} - J\mathbf{s})) \quad \mathbf{s} \in \mathbb{R}^M, \mathbf{u} \in \mathbb{R}^M, \tag{22}$$

To obtain Q_G, we observe that for fixed $\mathbf{H}$, the distribution of $\mathbf{U}$ given $\mathbf{Y}$ is also Gaussian.

$$\begin{aligned}
\mathbf{m}_{U \mid Y,H}(\mathbf{y}, \mathbf{H}) &= \mathrm{E}[\mathbf{U} \mid H = \mathbf{H}] + \mathrm{Cov}(\mathbf{U}, \mathbf{Y} \mid H = \mathbf{H}) \cdot \mathrm{Cov}(Y \mid H = \mathbf{H})^{-1} \cdot (\mathbf{y} - \mathrm{E}[\mathbf{y} \mid H = \mathbf{H}]) \\
\Sigma_{U \mid Y,H}(\mathbf{H}) &= \mathrm{Cov}(\mathbf{U} \mid H = \mathbf{H}) - \mathrm{Cov}(\mathbf{U}, \mathbf{Y} \mid H = \mathbf{H}) \cdot \mathrm{Cov}(Y \mid H = \mathbf{H})^{-1} \cdot \mathrm{Cov}(\mathbf{Y}, \mathbf{U} \mid H = \mathbf{H})
\end{aligned}$$

We now claim that $\Sigma_{U \mid Y,H}(\mathbf{H})$ is also nonsingular. This will be shown by proving that $\Sigma_{U,Y|H}(\mathbf{H})$ is positive definite, i.e.

$$(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)\Sigma_{U,Y|H}(\mathbf{H}) \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \mathrm{E}\left\{ \left(\boldsymbol{\alpha}^T \mathbf{U} + \boldsymbol{\beta}^T \mathbf{Y} \right)^2 \mid H = \mathbf{H} \right\} > 0 \quad \forall (\boldsymbol{\alpha}, \boldsymbol{\beta}) \neq \mathbf{0} \tag{23}$$

Now, by (6) and (9),

$$\mathbf{Y} = H(-\mathbf{F} + \mathbf{I})\mathbf{S} + H\mathbf{U} + \mathbf{Z}$$

By our second regularity assumption, the covariance of $(\mathbf{U}, \mathbf{S})$ is nonsingular. It follows that Σ_U is positive definite. We thus conclude that (23) holds for $\beta = 0$. If, on the other hand $\beta \neq \mathbf{0}$, then

$$\mathrm{E}\left\{ \left(\boldsymbol{\alpha}^T \mathbf{U} + \boldsymbol{\beta}^T \mathbf{Y} \right)^2 \mid H = \mathbf{H} \right\} = \mathrm{E}\left\{ \left(\boldsymbol{\alpha}^T \mathbf{U} + \boldsymbol{\beta}^T \left(H(-\mathbf{F} + \mathbf{I})\mathbf{S} + H\mathbf{U} \right) \right)^2 \mid H = \mathbf{H} \right\} + \mathrm{E}\left\{ \left(\boldsymbol{\beta}^T \mathbf{Z} \right)^2 \right\} > 0$$

since Z is independent of $\mathbf{X}$, $\mathbf{S}$ and H, and its covariance, Σ_Z, is nonsingular. This proves our claim.

Using similar arguments as in the above development of q_G, we obtain

$$\mathbf{m}_{U \mid Y,H}(\mathbf{y}, \mathbf{H}) = K(\mathbf{H})\mathbf{y}$$

where,

$$K(\mathbf{H}) = \left[(\mathbf{F}\Sigma_S + \mathbf{F}\Sigma_{S,X} + \Sigma_{S,X}^T + \Sigma_X)\mathbf{H}^T \right] \left[\mathbf{H}(\Sigma_S + \Sigma_X + \Sigma_{S,X} + \Sigma_{S,X}^T)\mathbf{H}^T + \Sigma_Z \right]^{-1}$$

and,

$$\begin{aligned}
\Sigma_{U \mid Y,H}(\mathbf{H}) = (\mathbf{F}\Sigma_S\mathbf{F}^T + \mathbf{F}\Sigma_{S,X} + \Sigma_{S,X}^T\mathbf{F}^T + \Sigma_X) - \\
\left[(\mathbf{F}\Sigma_S + \mathbf{F}\Sigma_{S,X} + \Sigma_{S,X}^T + \Sigma_X)\mathbf{H}^T \right] \left[\mathbf{H}(\Sigma_S + \Sigma_X + \Sigma_{S,X} + \Sigma_{S,X}^T)\mathbf{H}^T + \Sigma_Z \right]^{-1} \times \\
\left[(\mathbf{F}\Sigma_S + \mathbf{F}\Sigma_{S,X} + \Sigma_{S,X}^T + \Sigma_X)\mathbf{H}^T \right]^T
\end{aligned} \tag{24}$$

Observe that $K(\mathbf{H})$ and $\Sigma_{U \mid Y,H}(\mathbf{H})$ are fixed matrix functions of the matrices that constitute the problem setting, and of $\mathbf{H}$. Hence,

$$\begin{aligned}
Q_G(\mathbf{u} \mid \mathbf{y}, \mathbf{H}) &= \frac{1}{\sqrt{\det(2\pi\Sigma_{U \mid Y,H}(\mathbf{H}))}} \exp(-\frac{1}{2}(\mathbf{u} - K(\mathbf{H})\mathbf{y})^T \Sigma_{U \mid Y,H}(\mathbf{H})^{-1}(\mathbf{u} - K(\mathbf{H})\mathbf{y})) \\
&\qquad \mathbf{y} \in \mathbb{R}^N, \mathbf{H} \in \mathcal{R}_H, \mathbf{u} \in \mathbb{R}^M
\end{aligned} \tag{25}$$

We observe that $q_G(\mathbf{u} \mid \mathbf{s})$ is positive-valued for all $\mathbf{u} \in \mathbb{R}^M$ and $\mathbf{s} \in \mathbb{R}^M$. Similarly, $Q_G(\mathbf{u} \mid \mathbf{y}, \mathbf{H})$ is positive-valued for all $\mathbf{u} \in \mathbb{R}^M$, $\mathbf{y} \in \mathbb{R}^N$ and $\mathbf{H} \in \mathcal{R}_H$, where $\mathcal{R}_H$ is the support region of H. Therefore, they satisfy this condition of Lemma 1. The conditions of Lemma 1 also require that Q_G be the marginal distribution of $\mathbf{U}$ given $\mathbf{y}$ and $\mathbf{H}$, when the distribution of $\mathbf{U}$ is determined from the densities $f(\mathbf{s})$ and $q_G(\mathbf{u} \mid \mathbf{s})$. This is satisfied by definition.

We proceed by showing that the two functions q_G and Q_G admit Lagrange multipliers. Finding a Lagrange multiplier $\beta(\mathbf{y}, \mathbf{H})$ to satisfy (18) is easy. As in the discussion following (47), we have

$$\int_{\mathbf{s}\in\mathbb{R}^M} f(\mathbf{s}) f(\mathbf{y}, \mathbf{H} \mid \mathbf{s}, \mathbf{x} = \mathbf{f}(\mathbf{u}, \mathbf{s})) \frac{q_G(\mathbf{u} \mid \mathbf{s})}{Q_G(\mathbf{u} \mid \mathbf{y}, \mathbf{H})}\, d\mathbf{s} = q_G(\mathbf{y}, \mathbf{H}) \qquad \forall \mathbf{u} \in \mathbb{R}^M, \forall \mathbf{y} \in \mathbb{R}^N, \forall \mathbf{H} \in \mathcal{R}_H$$

Thus, defining $\beta(\mathbf{y}, \mathbf{H}) = -q_G(\mathbf{y}, \mathbf{H})$, (18) is satisfied.

We now turn our attention to the other Lagrange multipliers and to (17). Let $\mathbf{u}$ and $\mathbf{s}$ be fixed and let $\mathbf{x} \triangleq \mathbf{f}(\mathbf{u}, \mathbf{s})$. Simple manipulations of (17) lead to,

$$\begin{aligned}
&\int_{\mathbf{y}\in\mathbb{R}^N} \int_{\mathbf{H}\in\mathcal{R}_H} f(\mathbf{y}, \mathbf{H} \mid \mathbf{s}, \mathbf{x}) \log Q_G(\mathbf{u} \mid \mathbf{y}, \mathbf{H})\, d\mathbf{H}\, d\mathbf{y} - \\
&\qquad - \int_{\mathbf{y}\in\mathbb{R}^N} \int_{\mathbf{H}\in\mathcal{R}_H} f(\mathbf{y}, \mathbf{H} \mid \mathbf{s}, \mathbf{x})\, d\mathbf{H}\, d\mathbf{y} \cdot [\log q_G(\mathbf{u} \mid \mathbf{s}) + 1] + \\
&\qquad + < \Upsilon, \mathbf{x} \cdot \mathbf{x}^T > + < \Gamma, \mathbf{s} \cdot \mathbf{x}^T > + \frac{\alpha(\mathbf{s})}{f(\mathbf{s})} = 0
\end{aligned}$$

We continue,

$$\begin{aligned}
&\int_{\mathbf{y}\in\mathbb{R}^N} \int_{\mathbf{H}\in\mathcal{R}_H} f(\mathbf{y}, \mathbf{H} \mid \mathbf{s}, \mathbf{x}) \log Q_G(\mathbf{u} \mid \mathbf{y}, \mathbf{H})\, d\mathbf{H}\, d\mathbf{y} - \log q_G(\mathbf{u} \mid \mathbf{s}) + < \Upsilon, \mathbf{x} \cdot \mathbf{x}^T > + < \Gamma, \mathbf{s} \cdot \mathbf{x}^T > \\
&\qquad + \left[\frac{\alpha(\mathbf{s})}{f(\mathbf{s})} - 1\right] = 0
\end{aligned} \tag{26}$$

We begin by examining the first element in the above sum. This element is equal to,

$$\begin{aligned}
&\mathrm{E}_{Y,H}\left[\log Q_G(\mathbf{u} \mid \mathbf{Y}, H) \mid \mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}\right] = \\
&\quad = -\frac{1}{2}\mathrm{E}_{Y,H}\left[\log\det(2\pi\Sigma_{U \mid Y,H}(H)) \mid \mathbf{x}, \mathbf{s}\right] \\
&\qquad -\frac{1}{2}\mathrm{E}_{Y,H}\left[(\mathbf{u} - K(H)\mathbf{Y})^T \Sigma_{U \mid Y,H}(H)^{-1}(\mathbf{u} - K(H)\mathbf{Y}) \mid \mathbf{x}, \mathbf{s}\right] \\
&\quad = -\frac{1}{2}\mathrm{E}_{H}\left[\log\det(2\pi\Sigma_{U \mid Y,H}(H))\right] \\
&\qquad -\frac{1}{2}\mathrm{E}_{H}\left\{\mathrm{E}_{Y}\left[(\mathbf{u} - K(H)\mathbf{Y})^T \Sigma_{U \mid Y,H}(H)^{-1}(\mathbf{u} - K(H)\mathbf{Y}) \mid \mathbf{x}, \mathbf{s}, H\right]\right\}
\end{aligned} \tag{27}$$

We now focus on the contents of the braces. We use $\mathbf{u} = \mathbf{F}\mathbf{s} + \mathbf{x}$, $\mathbf{Y} = \mathbf{H}(\mathbf{x} + \mathbf{s}) + \mathbf{Z}$ to obtain,

$$\begin{aligned}
&\mathrm{E}_{Y}\left[(\mathbf{u} - K(\mathbf{H})\mathbf{Y})^T \Sigma_{U \mid Y,H}(\mathbf{H})^{-1}(\mathbf{u} - K(\mathbf{H})\mathbf{Y}) \mid \mathbf{x}, \mathbf{s}, \mathbf{H}\right] = \\
&\quad \mathbf{x}^T\left[(I - K(\mathbf{H})\mathbf{H})^T \Sigma_{U \mid Y,H}(\mathbf{H})^{-1}(I - K(\mathbf{H})\mathbf{H})\right]\mathbf{x} + \\
&\quad + \mathbf{s}^T\left[(\mathbf{F} - K(\mathbf{H})\mathbf{H})^T \Sigma_{U \mid Y,H}(\mathbf{H})^{-1}(\mathbf{F} - K(\mathbf{H})\mathbf{H})\right]\mathbf{s} + \\
&\quad 2\mathbf{s}^T\left[(\mathbf{F} - K(\mathbf{H})\mathbf{H})^T \Sigma_{U \mid Y,H}(\mathbf{H})^{-1}(I - K(\mathbf{H})\mathbf{H})\right]\mathbf{x} + \\
&\quad + \mathrm{tr}\left[K(\mathbf{H})^T \Sigma_{U \mid Y,H}(\mathbf{H})^{-1} K(\mathbf{H}) + \Sigma_Z\right]
\end{aligned}$$

Thus, we can rewrite (27) as,

$$\mathbf{x}^T A\mathbf{x} + \mathbf{s}^T B\mathbf{s} + \mathbf{s}^T C\mathbf{x} + D = < A, \mathbf{x}\cdot\mathbf{x}^T > + < B, \mathbf{s}\cdot\mathbf{s}^T > + < C, \mathbf{s}\cdot\mathbf{x}^T > + D \tag{28}$$

where,

$$\begin{aligned}
A &= -\frac{1}{2}E_H\left[(I - K(H)H)^T \Sigma_{U\,|\,Y,H}(H)^{-1}(I - K(H)H)\right] && (29)\\
B &= -\frac{1}{2}E_H\left[(\mathbf{F} - K(H)H)^T \Sigma_{U\,|\,Y,H}(H)^{-1}(\mathbf{F} - K(H)H)\right] && (30)\\
C &= -E_H\left[(\mathbf{F} - K(H)H)^T \Sigma_{U\,|\,Y,H}(H)^{-1}(I - K(H)H)\right] && (31)\\
D &= -\frac{1}{2}\mathrm{E}_H\left[\log\det(2\pi\Sigma_{U\,|\,Y,H}(H))\right] - \frac{1}{2}E_H\left\{\mathrm{tr}\left[K(H)^T\Sigma_{U\,|\,Y,H}(H)^{-1}K(H) + \Sigma_Z\right]\right\} && (32)
\end{aligned}$$

By the conditions of Theorem 1, the above expectations exist and are finite. Turning to the second element of the sum in (26) we obtain, using (22)

$$-\log q_G(\mathbf{u}\,|\,\mathbf{s}) = \frac{1}{2}\log\det(2\pi\Sigma_{U\,|\,S}) + \frac{1}{2}(\mathbf{u} - J\mathbf{s})^T\Sigma_{U\,|\,S}^{-1}(\mathbf{u} - J\mathbf{s}) \tag{33}$$

Applying a similar development to that of (27), we can rewrite (33) as,

$$< \hat{A}, \mathbf{x}\cdot\mathbf{x}^T > + < \hat{B}, \mathbf{s}\cdot\mathbf{s}^T > + < \hat{C}, \mathbf{s}\cdot\mathbf{x}^T > + \hat{D} \tag{34}$$

where

$$\begin{aligned}
\hat{A} &= \frac{1}{2}\Sigma_{U\,|\,S}^{-1}\\
\hat{B} &= \frac{1}{2}(\mathbf{F} - J)^T\Sigma_{U\,|\,S}^{-1}(\mathbf{F} - J)\\
\hat{C} &= (\mathbf{F} - J)^T\Sigma_{U\,|\,S}^{-1}\\
\hat{D} &= \frac{1}{2}\log\det(2\pi\Sigma_{U\,|\,S})
\end{aligned}$$

Using (28) and (34), we can rewrite (26) as,

$$< A + \hat{A} + \Upsilon, \mathbf{x}\cdot\mathbf{x}^T > + < B + \hat{B}, \mathbf{s}\cdot\mathbf{s}^T > + < C + \hat{C} + \Gamma, \mathbf{s}\cdot\mathbf{x}^T > + D + \hat{D} + \left[\frac{\alpha(\mathbf{s})}{f(\mathbf{s})} - 1\right] = 0$$

Finally, we may select our Lagrange multipliers for (17) as follows, completing the proof of Theorem 1.

$$\Upsilon = -(A + \hat{A}), \quad \Gamma = -(C + \hat{C}), \quad \alpha(\mathbf{s}) = f(\mathbf{s})\left[1 - D - \hat{D} - < B + \hat{B}, \mathbf{s}\cdot\mathbf{s}^T >\right]$$

$\square$

Note that with linear-assignment, when $\mathbf{X}$ and $\mathbf{S}$ are jointly-Gaussian, the achievable rate $I(\mathbf{U};\mathbf{Y},H) - I(\mathbf{U};\mathbf{S})$ is a function of the setting (as defined in Definition 1). The expression for the achievable rate can be computed as follows,

$$I(\mathbf{U};\mathbf{Y},H) - I(\mathbf{U};\mathbf{S}) = h(\mathbf{U}\,|\,\mathbf{S}) - h(\mathbf{U}\,|\,\mathbf{Y},H) = \frac{1}{2}\log\det\Sigma_{U\,|\,S} - \frac{1}{2}E_H\left[\log\det\Sigma_{U\,|\,Y,H}(H)\right] \tag{35}$$

The last equation is obtained from the following discussion. For fixed $\mathbf{s}$, the marginal distribution of $\mathbf{U}$ given $\mathbf{S} = \mathbf{s}$ is zero-mean Gaussian distributed with variance $\Sigma_{U\,|\,S}$ (which is given by (21) and is independent

of $\mathbf{s}$). For fixed $\mathbf{y}$ and $\mathbf{H}$, the marginal distribution of $\mathbf{U}$ given $\mathbf{Y} = \mathbf{y}$ and $H = \mathbf{H}$ is zero-mean Gaussian distributed with variance $\Sigma_{U \mid Y,H}(\mathbf{H})$ (which is given by (24) and is independent of $\mathbf{y}$ but dependent on $\mathbf{H}$).

Note that the achievability proof of Gel'fand and Pinsker [14], that states that we may indeed achieve the rate $I(\mathbf{U};\mathbf{Y};H) - I(\mathbf{U};\mathbf{S})$ assumes that the random variables involved are discrete-valued. In Appendix IV we use quantization arguments to prove that $F(q_G, Q_G)$, defined using (11) (which assumes *continuous* random variables), is indeed achievable.

IV. The Linear-Assignment Fading-Paper (LAFP) Achievable Region

A. Definition

In Sec. II-D we described how dirty-paper transmission methods can be used to construct an algorithm for transmission over the non-fading MIMO-BC channel. The same approach can be used to construct an algorithm for transmission over the fading MIMO-BC channel, using the linear-assignment fading-paper transmission methods of Sec. III.

In our approach, we rely on Theorem 1 and confine our attention to Gaussian distributions for the signals $\{\mathbf{X}_l\}_{l=1}^L$, defined as in Sec. II-D. Our choice is greedy in the sense that we seek to maximize the rate to each user individually, while a global perspective could possibly prescribe a different choice. However, a similar choice in the definition of the dirty-paper achievable region was eventually proven to coincide with the global optimum as well. We refer to the convex-hull of the union of rate regions that are achievable using this approach, as the *linear-assignment fading-paper* (LAFP) achievable region.

The analysis of Weingarten $et\ al.$ [30] does not apply to the fading setting. Furthermore, linear-assignments have not been proven to exhaust the capacity of the fading-paper channel. Thus, unlike the dirty-paper achievable region of Sec. II-D, the LAFP achievable region is not guaranteed to be optimal.

The determination of the dirty paper achievable region of Sec. II-D involves determining the covariance matrices $\Sigma_X^{(l)}$ for the various signals $\mathbf{X}_l$ (see e.g. [6] and [29]). However, each signal $\mathbf{X}_l$ is assumed to be independent of the interference $\mathbf{S}_l \triangleq \sum_{i<l} \mathbf{X}_i$, and Gaussian. In our above definition of the LAFP, we have not restricted ourselves to signals $\{\mathbf{X}_l\}_{l=1}^L$ that are independent of their respective interferences $\{\mathbf{S}_l\}_{l=1}^L$. Thus, in addition to determining $\Sigma_X^{(l)}$, it would appear that we must determine the covariance $\Sigma_{X,S}^{(l)}$, between $\mathbf{X}_l$ and $\mathbf{S}_l$ as well.

However, the following theorem proves that we may indeed confine ourselves to $\Sigma_{X,S}^{(l)} = \mathbf{0}$, without loss of optimality.

Theorem 2: The LAFP achievable region is exhausted by a choice of random variables $\{\mathbf{X}_l\}_{l=1}^L$ for the various users that are independent of their respective interferences $\{\mathbf{S}_l\}_{l=1}^L$

The proof of this theorem is provided in Appendix V.

Note that in this theorem we do *not* claim that for the given fading-paper problem observed by user l, selecting $\mathbf{X}_l$ to be independent of $\mathbf{S}_l$ incurs no loss of optimality. Rather, the proof involves replacing an

entire given set of signals $\mathbf{X}_1, ..., \mathbf{X}_L$, which may not be independent (corresponding to some set of achievable rates on the LAFP achievable region) with a new set $\hat{\mathbf{X}}_1, ..., \hat{\mathbf{X}}_L$ that are independent, without sacrificing the rates of the individual users. In the resulting set, user l's signal $\hat{\mathbf{X}}_l$ is indeed independent of $\hat{\mathbf{S}}_l = \sum_{i<l} \hat{\mathbf{X}}_i$. However, the independence was achieved also by altering the fading-paper problem this user faces.

B. Comparison with Dirty-Paper Transmission

So far, we have focused on similarities between the dirty-paper transmission over a fixed MIMO-BC and LAFP transmission over a fading MIMO-BC channel. Both approaches use linear strategies, both employ independently distributed Gaussian random variables to construct their signals to the receivers.

However, the two methods differ in two important ways.

1) The choice of the constant matrix $\mathbf{F}$ in dirty-paper transmission is based on the fixed channel matrix $\mathbf{H}$. With fading-paper, only the statistics of $\mathbf{H}$ are known and thus $\mathbf{F}$ must be selected differently.
2) The fading-paper receiver accounts for a channel fade $\mathbf{H}$ that fluctuates from one time instance to another. The dirty-paper receiver assumes that $\mathbf{H}$ is fixed. More precisely, the dirty paper decoder seeks a codeword that is jointly typical with $\mathbf{y}$, while the fading paper decoder seeks a codeword that is jointly typical with both $\mathbf{y}$ and $\mathbf{H}$.

Despite these two shortcomings, dirty-paper transmission can still be applied to a fading-paper channel by simply assuming that $\mathbf{H}$ is fixed at its average, and treating its fluctuations as noise. For a fading paper transmission strategy to be interesting, we must demonstrate that its performance surpasses that of dirty-paper transmission.

An evaluation of the dirty-paper achievable region (i.e., when the transmitter and receiver assume that the channel is fixed at its expected value $\mathrm{E}H$) over the fading MIMO-BC scheme is difficult. This is because of the operation of the decoder, which uses a mismatched model of the channel. However, we may obtain an outer bound on the dirty-paper achievable region if we replace the receiver with an optimal LAFP receiver that uses the channel information available to it (unlike the standard dirty-paper receiver). In this case, the achievable rate may be obtained from (35). With the dirty-paper achievable region, however, the matrices $\mathbf{F}$ (for each instance of Σ_X, Σ_S and Σ_Z for the user) are not the optimal fading paper matrices, but rather are computed using (5), under the assumption of a fixed channel matrix, equal to $\mathrm{E}H$. Under these conditions, the approach differs from LAFP only in the way the matrix $\mathbf{F}$ is selected.

We let $\mathbf{F}_{DPC}(\mathbf{H})$ denote the choice of $\mathbf{F}$ with dirty-paper transmission over a channel whose fixed channel matrix is $\mathbf{H}$. That is, $\mathbf{F}_{DPC}(\mathbf{H})$ is a matrix function of $\mathbf{H}$, given by the right hand side of (5) (for brevity of notation, we neglect the reliance of $F_{DPC}(\cdot)$ on Σ_X and Σ_Z). With this notation, the choice of $\mathbf{F}$ that is used in the above-mentioned dirty paper like transmission strategy is $\mathbf{F}_{DPC}(\mathrm{E}H)$.

Evaluating the LAFP region involves determining the union of the regions obtained for all matrices $\mathbf{F}$. Equivalently, it involves maximizing (35) over $\mathbf{F}$ (e.g. using a grid search) given the covariances of X and S

(note that by Theorem 2 we set $\Sigma_{S,X} = 0$). However, we obtained an *inner* bound by restricting our attention, for each Σ_X and Σ_Z to the set

$$\mathcal{F} \triangleq \{ \mathbf{F}_{DPC}(\mathbf{H}) \, : \, \mathbf{H} \in \mathcal{R}_H \} \tag{36}$$

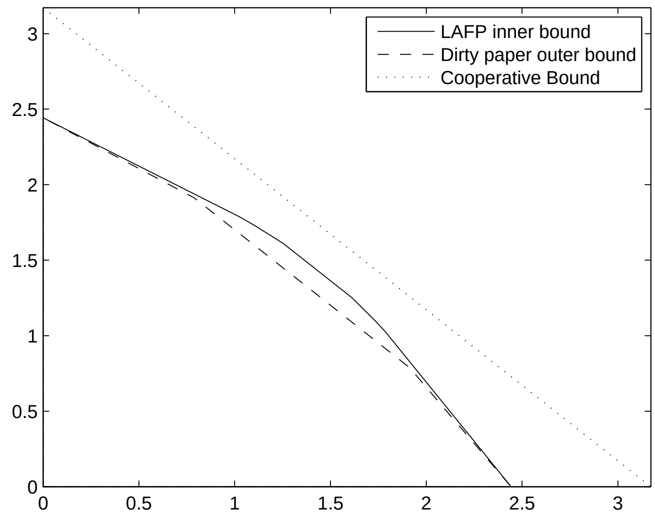

Fig. 1. Comparison between an inner bound on the LAFP achievable region and an outer bound on the dirty paper achievable region.

Fig. 1 presents a numerical example where the two approaches are compared. In this example, there are two users (receivers). The transmitter has two antennas ($M = 2$) and the receivers have one antenna each ($N_1 = N_2 = 1$). The power constraint is $P_{TOT} = 10$. The distributions of the channel matrices are given by,

$$H^{(1)} = \begin{cases} [1, 0.4] & \text{with probability } 1/2 \\ [1, 3] & \text{with probability } 1/2 \end{cases} \qquad H^{(2)} = \begin{cases} [0.4, 1] & \text{with probability } 1/2 \\ [3, 1] & \text{with probability } 1/2 \end{cases}$$

The noise variance at each receiver is 1.

The achievable regions in both cases (i.e. LAFP and dirty-paper) were found by first applying a grid search for the matrices $\Sigma_X^{(1)}$ and $\Sigma_X^{(2)}$. In line with Theorem 2, we assumed without loss of optimality that the two signals $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are independent.

For each such pair $\Sigma_X^{(1)}$ and $\Sigma_X^{(2)}$, the matrix $\mathbf{F}$ for user 2 was computed as described above. That is, for the LAFP achievable region, $\mathbf{F}$ was found by maximizing the achievable rate of user 2 over the set $\mathcal{F}$ (which is a function of the user's covariance matrix[9] $\Sigma_X^{(2)}$). For the dirty-paper achievable region, $\mathbf{F}_{DPC}(\mathrm{E}H)$ was used.

With both schemes, for fixed matrices $\Sigma_X^{(1)}$, $\Sigma_X^{(2)}$ and $\mathbf{F}$, the achievable rates R_1 and R_2 for the two users were computed as follows. R_1 was obtained using the following expression (recall that user 1's observed

[9]In the general case, where there are more than two users, $\mathcal{F}$ is also a function of $\sum_{l>2} \Sigma_X^{(l)}$, the unknown interference from subsequent users, which must be accounted for in the effective noise as explained in Appendix VI.

signal $Y^{(1)}$ is scalar in this example),

$$R_1 = \frac{1}{2} E_{H^{(1)}} \log \left(1 + \frac{H^{(1)} \Sigma_X^{(1)} H^{(1)T}}{H^{(1)} \Sigma_X^{(2)} H^{(1)T} + 1} \right)$$

R_2 is given by the right hand side of (35). Since we have assumed $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ to be independent, the expressions for $\Sigma_{U \mid S}$ and $\Sigma_{U \mid Y,H}(\mathbf{H})$ (which appear in (35)) are simple[10]. That is, $\Sigma_{U \mid S} = \Sigma_X^{(2)}$ and $\Sigma_{U \mid Y,H}(\mathbf{H})$ is obtained from (24) by setting $\Sigma_{S,X}$ to zero.

The maximal sum-rate on the dirty paper outer bound was 2.7 bits per channel use, while the maximum sum-rate on the LAFP inner bound was 2.86. This achievable rate was obtained by selecting,

$$\Sigma_X^{(1)} = \begin{bmatrix} 1 & 2-\epsilon \\ 2-\epsilon & 4 \end{bmatrix}, \quad \Sigma_X^{(2)} = \begin{bmatrix} 4.5 & -1.5+\epsilon \\ -1.5+\epsilon & 0.5 \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 1.0909 & 0.3636 \\ -0.3636 & -0.1212 \end{bmatrix}$$

where $0 < \epsilon \to 0$ such that $\Sigma_X^{(1)}$ and $\Sigma_X^{(2)}$ are positive definite. Thus, a simple approach, which uses knowledge of the channel distribution at the transmitter, was able to produce *at least* a 6% increase in throughput.

Although we have not established the optimality of the LAFP achievable region, we can obtain an idea of how far we are from the optimum using a cooperative upper bound on the achievable sum-capacity (i.e., the maximum achievable sum rate to all users), as suggested by Sato [23]. The use of such a bound in the context of the (non-fading) MIMO-BC channel was first suggested by Caire and Shamai [6]. Computation of cooperative upper-bounds for the above fading MIMO-BC example is discussed in Appendix VII. We obtained a bound of 3.17 on the maximum achievable sum-rate. Thus, in terms of the sum-rate, LAFP is capable of transmission at rates that are 10% below the optimum.

In Appendix VI we will discuss the computation of the LAFP achievable region with more than two users.

V. CONCLUSION

A. Suggestions for Further Research

1) **Heuristic methods for computing F.** Expression (36), with which we computed the matrix $\mathbf{F}$ for the LAFP region in Sec. IV-B, was developed heuristically. A different expression could possibly produce a substantially larger achievable region. One option would be to search for $\mathbf{F}$ along a fine grid (as noted in Sec. IV-B). An alternative option would be to apply a gradient ascent method, using $\mathbf{F}$ as defined in (36) as a starting point.
2) **A wider range of strategies.** The confinement to linear assignments as defined in Sec. III is in no way known to be optimal. Dupuis $et\ al.$ [11] suggested an algorithm that is based on the concepts of the Blahut-Arimoto algorithm, that can theoretically be used to evaluate the capacity of a general side-information channel (of which the fading paper channel is an instance). In practice, applying the algorithm requires evaluations over a set of *strategies* which is impossibly large. However, applying

[10]In the context of our discussion, $\mathbf{S} = \mathbf{X}^{(1)}$, $\mathbf{X} = \mathbf{X}^{(2)}$ and $\mathbf{Z}$ has covariance $\Sigma_Z = 1$. $\mathbf{U} = \mathbf{F}\mathbf{S} + \mathbf{X}$, as usual.

the algorithm over any subset of these strategies produces an achievable rate. This achievable rate may further narrow the gap to the cooperative upper-bound (as discussed in Sec. IV-B).

B. Concluding Remarks

The problem of transmitting over fading MIMO-BC channels is of great practical interest. In this paper we presented an achievable region for this channel that relies on fading-paper transmission strategies. Our main contribution is Theorem 1, which proves that a Gaussian distribution achieves the linear-assignment capacity. We believe that the approach we developed in the proof of that theorem, which employs convex-analysis methods, could be useful in further analysis of this channel.

In Sec. IV-B we have shown that a simple approach, which makes use of the channel distribution information available to the transmitter, easily produces a gain over dirty-paper transmission. Further research (perhaps in the lines of Sec. V-A) could produce further performance gains.

APPENDIX I
THE OPTIMAL ACHIEVABLE RATE WITH ZERO CHANNEL STATE INFORMATION AT THE TRANSMITTER

Consider a broadcast channel where all the receivers have the same number of antennas. We wish to show that capacity in this case is achieved by time-sharing among the users.

A channel model that assumes zero knowledge of the channel fade to each of the users, effectively assumes that all channels are the same. The signals at the different receivers are equivalent in their statistical properties, and thus each receiver is capable, beside decoding its own signal, of decoding all the messages to the other users as well. Thus, the sum-rate of this system is upper-bounded by the single-user rate of each of the users. Such a capacity region is exhausted by time-sharing.

APPENDIX II
THE OPTIMAL MATRIX $\mathbf{F}$ IN THE ACHIEVABILITY PROOF FOR DIRTY-PAPER

In this appendix we prove the optimality of $\mathbf{F}$ as defined by (5). We let $\mathbf{U}$ and $\mathbf{X}$ be defined as in the discussion preceding (5). The achievable rate with this choice is given by $I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U},\mathbf{S})$ (see (4)). We now seek to prove that this rate coincides with the capacity of the corresponding no-interference channel defined by (3). Our proof follows in the lines of a similar proof by Cohen and Lapidoth [6] for the scalar dirty-paper channel.

To obtain our result, we prove a stronger result. We prove that for any choice of Σ_X, letting $\mathbf{F}$ be given by (5), we obtain that the achievable rate coincides with the achievable rate $I(\mathbf{X};\hat{\mathbf{Y}})$ for the no-interference channel (3).

Our objective is to show that the achievable rate $I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U},\mathbf{S})$, with this choice of $\mathbf{F}$, coincides with the achievable rate of the no-interference channel when the input $\mathbf{X}$ is distributed as $\mathcal{N}(\mathbf{0},\Sigma_X)$.

Let $\hat{\mathbf{X}} = \mathbf{W}\hat{\mathbf{Y}}$ be the linear minimum mean-square error (LMMSE) estimate for $\mathbf{X}$ given $\hat{\mathbf{Y}}$. $\mathbf{W}$ is obtained by [19],

$$\mathbf{W} = \mathrm{Cov}(\mathbf{X}, \hat{\mathbf{Y}})\mathrm{Cov}(\hat{\mathbf{Y}})^{-1} = \Sigma_X \mathbf{H}^T (\mathbf{H}\Sigma_X \mathbf{H}^T + \Sigma_Z)^{-1} \tag{37}$$

By definition of the LMMSE estimate, the error $\mathbf{E} \triangleq \mathbf{X} - \hat{\mathbf{X}}$ is uncorrelated with $\hat{\mathbf{Y}}$. Since $\mathbf{E}$ and $\hat{\mathbf{Y}}$ are jointly-Gaussian, they are also independent. $\mathbf{S}$ is independent of both, and thus $\mathbf{E}$ is independent of $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{HS}$.

Examining $I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U},\mathbf{S})$, we have

$$I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U},\mathbf{S}) = h(\mathbf{U} \mid \mathbf{S}) - h(\mathbf{U} \mid \mathbf{Y}) \tag{38}$$

We now examine both elements of the difference on the right hand side of the above.

$$h(\mathbf{U} \mid \mathbf{S}) = h(\mathbf{FS} + \mathbf{X} \mid \mathbf{S}) = h(\mathbf{X}) \tag{39}$$

where the last equation is obtained by the fact that $\mathbf{S}$ and $\mathbf{X}$ are independent.

$$\begin{aligned}
h(\mathbf{U} \mid \mathbf{Y}) &= h(\mathbf{FS} + \mathbf{X} \mid \mathbf{Y}) \overset{(a)}{=} h(\mathbf{WHS} + \mathbf{X} \mid \mathbf{Y}) \\
&= h(\mathbf{WHS} + \mathbf{X} - \mathbf{WY} \mid \mathbf{Y}) = h(\mathbf{WHS} + \mathbf{X} - \mathbf{W}(\mathbf{HS} + \mathbf{HX} + \mathbf{Z}) \mid \mathbf{Y}) \\
&= h(\mathbf{X} - \mathbf{W}(\mathbf{HX} + \mathbf{Z}) \mid \mathbf{Y}) = h(\mathbf{X} - \hat{\mathbf{X}} \mid \mathbf{Y}) = h(\mathbf{E} \mid \mathbf{Y}) \overset{(b)}{=} h(\mathbf{E}) \overset{(c)}{=} h(\mathbf{E} \mid \hat{\mathbf{Y}}) \\
&= h(\mathbf{X} - \mathbf{W}\hat{\mathbf{Y}} \mid \hat{\mathbf{Y}}) = h(\mathbf{X} \mid \hat{\mathbf{Y}})
\end{aligned} \tag{40}$$

Equality (a) is obtained from the observation that the right hand side of (5) equals $\mathbf{W} \cdot \mathbf{H}$ where $\mathbf{W}$ is given by (37). Equalities (b) and (c) are obtained from the fact that $\mathbf{E}$ is independent of $\hat{\mathbf{Y}}$ and $\mathbf{Y}$. Finally, combining (38), (39) and (40) we obtain our desired result,

$$I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U},\mathbf{S}) = h(\mathbf{X}) - h(\mathbf{X} \mid \hat{\mathbf{Y}}) = I(\mathbf{X};\hat{\mathbf{Y}})$$

□

APPENDIX III
PROOF OF LEMMA 1

Let q and Q be a pair of feasible functions for (12). We will now show that $F(q,Q) \le F(q^\star, Q^\star)$.

$$\begin{aligned}
F(q,Q) - F(q^\star, Q^\star) &= \int_{\mathbf{s}\in\mathbb{R}^M} \int_{\mathbf{u}\in\mathbb{R}^M} \int_{\mathbf{y}\in\mathbb{R}^N} \int_{\mathbf{H}\in\mathcal{R}_H} f(\mathbf{s}) f(\mathbf{y},\mathbf{H} \mid \mathbf{s}, \mathbf{x} = \mathbf{f}(\mathbf{u},\mathbf{s})) \cdot \\
&\quad \cdot \left[q(\mathbf{u} \mid \mathbf{s}) \log \frac{Q(\mathbf{u} \mid \mathbf{y}, \mathbf{H})}{q(\mathbf{u} \mid \mathbf{s})} - q^\star(\mathbf{u} \mid \mathbf{s}) \log \frac{Q^\star(\mathbf{u} \mid \mathbf{y}, \mathbf{H})}{q^\star(\mathbf{u} \mid \mathbf{s})} \right] d\mathbf{H}\, d\mathbf{y}\, d\mathbf{u}\, d\mathbf{s}
\end{aligned} \tag{41}$$

Let $l(x,y) \triangleq x \cdot \log(y/x)$. This function is jointly-concave in its arguments. By the gradient inequality [4][Chapter 3, Section 3.1.3] for concave functions, we have for arbitrary $x, y \in \mathbb{R}_+$ and $x^\star, y^\star \in \mathbb{R}_{++}$,

$$l(x,y) - l(x^\star, y^\star) \le l_x(x^\star, y^\star) \cdot (x - x^\star) + l_y(x^\star, y^\star) \cdot (y - y^\star)$$

where l_x and l_y denote the partial derivatives of l with respect to x and y, respectively. Thus, we can bound (41) by,

$$\begin{aligned} F(q,Q) - F(q^\star,Q^\star) &\leq \int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M}\int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\mid\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))\cdot \\ &\cdot\ [l_x(q^\star(\mathbf{u}\mid\mathbf{s}),Q^\star(\mathbf{u}\mid\mathbf{y},\mathbf{H}))\cdot(q(\mathbf{u}\mid\mathbf{s})-q^\star(\mathbf{u}\mid\mathbf{s}))+ \\ &+\ l_y(q^\star(\mathbf{u}\mid\mathbf{s}),Q^\star(\mathbf{u}\mid\mathbf{y},\mathbf{H}))\cdot(Q(\mathbf{u}\mid\mathbf{y},\mathbf{H})-Q^\star(\mathbf{u}\mid\mathbf{y},\mathbf{H}))]\,d\mathbf{H}\,d\mathbf{y}\,d\mathbf{u}\,d\mathbf{s} \end{aligned} \quad (42)$$

In the development below, we will show that this integral equals zero. This will then conclude the proof of the lemma.

To prove this, we will show that the two integrals below equal zero. For simplicity of notation, we let q and Q denote $q(\mathbf{u}\mid\mathbf{s})$ and $Q(\mathbf{u}\mid\mathbf{y},\mathbf{H})$, respectively.

$$\int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M}\int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\mid\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))\cdot l_x(q^\star,Q^\star)\cdot(q-q^\star)\,d\mathbf{H}\,d\mathbf{y}\,d\mathbf{u}\,d\mathbf{s} = 0 \quad (43)$$

$$\int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M}\int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\mid\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))\cdot l_y(q^\star,Q^\star)\cdot(Q-Q^\star)\,d\mathbf{H}\,d\mathbf{y}\,d\mathbf{u}\,d\mathbf{s} = 0 \quad (44)$$

We first prove (43). Multiplying (17) by $q-q^\star$, and using the fact that $l_x(x,y)=\log(y/x)-1$, we get

$$\begin{aligned} &\left[\int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\mid\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))l_x(q^\star,Q^\star)\,d\mathbf{H}\,d\mathbf{y}\right](q-q^\star)+ \\ &\quad\left[f(\mathbf{s})<\Upsilon,\mathbf{f}(\mathbf{u},\mathbf{s})\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T>\right](q-q^\star)+\left[f(\mathbf{s})<\Gamma,\mathbf{s}\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T>\right](q-q^\star)+\alpha(\mathbf{s})(q-q^\star)=0 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\mathbf{s}\in\mathbb{R}^M,\forall\mathbf{u}\in\mathbb{R}^M \end{aligned}$$

Integrating the above with respect to $\mathbf{u}$ and $\mathbf{s}$ would yield zero. We now focus on the integrals of the individual elements of the above sum. The first integral is equal to the left hand side of (43). To prove this integral is zero, we will show that the other integrals are zero. This will yield (43).

We first integrate with respect to $\mathbf{u}$ and then $\mathbf{s}$. The order of integration matters, because the range of the integration is unbounded, and some of the integrands are not non-negative and not necessarily Lebesgue-integrable (i.e., the integral of their absolute value may be infinite).

$$\begin{aligned} &\int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M}\left[f(\mathbf{s})<\Upsilon,\mathbf{f}(\mathbf{u},\mathbf{s})\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T>\right](q-q^\star)\,d\mathbf{u}\,d\mathbf{s} \\ &=<\Upsilon,\int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M} f(\mathbf{s})\left[\mathbf{f}(\mathbf{u},\mathbf{s})\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T\right]\cdot q\,d\mathbf{u}\,d\mathbf{s} \\ &-\int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M} f(\mathbf{s})\left[\mathbf{f}(\mathbf{u},\mathbf{s})\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T\right]\cdot q^\star\,d\mathbf{u}\,d\mathbf{s}> \\ &=<\Upsilon,\Sigma_X-\Sigma_X>=0 \end{aligned}$$

The equality before last results from (13) and from the feasibility of the functions q and $q^\star$. In a similar way, using (14), we obtain that,

$$\int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M}\left[f(\mathbf{s})<\Gamma,\mathbf{s}\cdot\mathbf{f}(\mathbf{u},\mathbf{s})^T>\right](q-q^\star)\,d\mathbf{u}\,d\mathbf{s}=0$$

Finally, we examine the last integral.

$$\int_{\mathbf{s}\in\mathbb{R}^M}\int_{\mathbf{u}\in\mathbb{R}^M}\alpha(\mathbf{s})(q-q^\star)\,d\mathbf{u}\,d\mathbf{s} = \int_{\mathbf{s}\in\mathbb{R}^M}\alpha(\mathbf{s})\left[\int_{\mathbf{u}\in\mathbb{R}^M} q\,d\mathbf{u} - \int_{\mathbf{u}\in\mathbb{R}^M} q^\star\,d\mathbf{u}\right]\,d\mathbf{s}$$
$$= \int_{\mathbf{s}\in\mathbb{R}^M}\alpha(\mathbf{s})\left[1-1\right]\,d\mathbf{s} = 0$$

The equality before last results from (15). Thus, we obtain (43).

Similarly, relying on (18) and (16), we obtain,

$$\int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H}\int_{\mathbf{u}\in\mathbb{R}^M}\int_{\mathbf{s}\in\mathbb{R}^M} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\mid\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))\cdot l_y(q^\star,Q^\star)\cdot(Q-Q^\star)\,d\mathbf{s}\,d\mathbf{u}\,d\mathbf{H}\,d\mathbf{y} = 0 \quad (45)$$

The order of integration, unfortunately, is not that of (44). To prove that we may change the order of integration, we must prove that the integrand is Lebesgue-integrable (Fubini's Theorem, see e.g. [2][Theorem 18.3]). To do this, we will prove that

$$\int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H}\int_{\mathbf{u}\in\mathbb{R}^M}\int_{\mathbf{s}\in\mathbb{R}^M} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\mid\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))\cdot l_y(q^\star,Q^\star)\cdot Q\,d\mathbf{s}\,d\mathbf{u}\,d\mathbf{H}\,d\mathbf{y} < \infty \quad (46)$$

Since the integrand in the above is nonnegative, this would yield that it is integrable. Since Q is arbitrary, the same would apply if we replace it with $Q^\star$. The integrand in (45), which is not necessarily nonnegative, is thus also integrable because it is obtained by subtracting the integrand in (46) by the same expression, with Q replaced by $Q^\star$.

Using $l_y(x,y) = x/y$, we may rewrite the left hand side of (46) as

$$\int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H}\int_{\mathbf{u}\in\mathbb{R}^M}\int_{\mathbf{s}\in\mathbb{R}^M} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\mid\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))\cdot\frac{q^\star}{Q^\star}\cdot Q\,d\mathbf{s}\,d\mathbf{u}\,d\mathbf{H}\,d\mathbf{y}$$
$$= \int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H}\int_{\mathbf{u}\in\mathbb{R}^M}\frac{Q}{Q^\star}\cdot\left[\int_{\mathbf{s}\in\mathbb{R}^M} f(\mathbf{s})f(\mathbf{y},\mathbf{H}\mid\mathbf{s},\mathbf{x}=\mathbf{f}(\mathbf{u},\mathbf{s}))\cdot q^\star\,d\mathbf{s}\right]\,d\mathbf{u}\,d\mathbf{H}\,d\mathbf{y} \quad (47)$$

The inside of the brackets is equal to $q^\star(\mathbf{y},\mathbf{H},\mathbf{u})$, defined to equal the marginal density of $\mathbf{Y}$, $\mathbf{H}$ and $\mathbf{U}$ where the distribution of $\mathbf{U}$ given $\mathbf{S}$ is determined by the density $q^\star$. Similarly defining $q^\star(\mathbf{y},\mathbf{H})$, we obtain by the conditions of Lemma 1, that $q^\star(\mathbf{y},\mathbf{H},\mathbf{u}) = q^\star(\mathbf{y},\mathbf{H})\cdot Q^\star(\mathbf{u}\mid\mathbf{y},\mathbf{H})$. Thus, (47) becomes,

$$\int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H}\int_{\mathbf{u}\in\mathbb{R}^M}\frac{Q}{Q^\star}\cdot q^\star(\mathbf{y},\mathbf{H})\cdot Q^\star\,d\mathbf{u}\,d\mathbf{H}\,d\mathbf{y} = \int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H} q^\star(\mathbf{y},\mathbf{H})\int_{\mathbf{u}\in\mathbb{R}^M} Q\,d\mathbf{u}\,d\mathbf{H}\,d\mathbf{y}$$
$$= \int_{\mathbf{y}\in\mathbb{R}^N}\int_{\mathbf{H}\in\mathcal{R}_H} q^\star(\mathbf{y},\mathbf{H})\cdot 1\,d\mathbf{H}\,d\mathbf{y} = 1 < \infty$$

Thus, by the above discussion, the order of integration in (45) can be changed, and we obtain (44). Coupled with (43), this proves that the right hand side of (42) is zero, concluding the proof of the lemma. □

APPENDIX IV
THE ACHIEVABILITY OF $F(q_G, Q_G)$

The random variables $\mathbf{U},\mathbf{S},\mathbf{Y},H$ that achieve the LAFP capacity are continuous. In practice one can only realize the Gelfand-Pinsker capacity of a set $\hat{\mathbf{U}},\hat{\mathbf{S}},\hat{\mathbf{Y}},\hat{H}$ of discrete random variables. We now show that $\mathbf{U},\mathbf{S},\mathbf{Y},H$ can be quantized to a set $\hat{\mathbf{U}},\hat{\mathbf{S}},\hat{\mathbf{Y}},\hat{H}$ of discrete random variables that can approach the LAFP

capacity arbitrarily close. The LAFP capacity is given by $R_{\text{achievable}} = F(q_G, Q_G)$ where $F(q, Q)$ is defined by (11).

We create a quantized version as follows. Let $\mathcal{B}_n(c, d)$ denote a cube in $\mathbb{R}^n$ with center c and size length d, i.e.,

$$\mathcal{B}_n(c, d) = \{(x_1, x_2, \ldots, x_n) \,:\, c - d/2 < x_i \le c + d/2,\ i = 1, \ldots, n\}$$

We define discrete random variables $\hat{\mathbf{S}}, \hat{\mathbf{U}}, \hat{\mathbf{Y}}, \hat{H}$ which are quantized versions of $\mathbf{S}, \mathbf{U}, \mathbf{Y}, H$, respectively, as follows. Recall that M and N are the dimensions of $\mathbf{S}$ and $\mathbf{Y}$, respectively. The dimension of H is thus $M \times N$. Fix some $\epsilon > 0$ sufficiently small, and $\rho > 0$ sufficiently large. Let $\mathbf{s}_i$, $i = 1, \ldots, N_s$ denote all the points in $\mathbb{R}^M$, such that $\mathbf{s}_i \in \mathcal{B}_M(0, \rho)$ and such that all the coordinates of $\mathbf{s}_i$ are integer multiples of ϵ. Similarly, let $\mathbf{u}_j$, $j = 1, \ldots, N_u$, $\mathbf{y}_k$, $k = 1, \ldots, N_y$ and $\mathbf{H}_l$, $l = 1, \ldots, N_h$ denote all the points in $\mathbb{R}^M$, $\mathbb{R}^N$ and $\mathcal{R}_H$, such that $\mathbf{u}_j \in \mathcal{B}_M(0, \rho)$, $\mathbf{y}_k \in \mathcal{B}_N(0, \rho)$ and $\mathbf{H}_l \in \mathcal{B}_{MN}(0, \rho)$, and such that all the coordinates of $\mathbf{u}_j$, $\mathbf{y}_k$ and $\mathbf{H}_l$ are integer multiples of ϵ.

We define by $\mathcal{S}_i$, $i = 0, 1, \ldots, N_s$ the following regions,

$$\mathcal{S}_i = \begin{cases} \mathbb{R}^M \bigcap \mathcal{B}_M(\mathbf{s}_i, \epsilon), & \text{if } i = 1, 2, \ldots, N_s; \\ \mathbb{R}^M \setminus \left[\bigcup_{i=1}^{N_s} \mathcal{B}_M(\mathbf{s}_i, \epsilon)\right], & \text{if } i = 0. \end{cases}$$

Similarly we define

$$\mathcal{U}_j = \begin{cases} \mathbb{R}^M \bigcap \mathcal{B}_M(\mathbf{u}_j, \epsilon), & \text{if } j = 1, 2, \ldots, N_u; \\ \mathbb{R}^M \setminus \left[\bigcup_{j=1}^{N_u} \mathcal{B}_M(\mathbf{u}_j, \epsilon)\right], & \text{if } j = 0. \end{cases}$$

$$\mathcal{Y}_k = \begin{cases} \mathcal{B}_N(\mathbf{y}_k, \epsilon), & \text{if } k = 1, 2, \ldots, N_y; \\ \mathbb{R}^N \setminus \bigcup_{k=1}^{N_y} \mathcal{B}_N(\mathbf{y}_k, \epsilon), & \text{if } k = 0. \end{cases}$$

and

$$\mathcal{H}_l = \begin{cases} \mathcal{R}_H \bigcap \mathcal{B}_{MN}(\mathbf{H}_l, \epsilon), & \text{if } l = 1, 2, \ldots, N_h; \\ \mathcal{R}_H \setminus \bigcup_{l=1}^{N_h} \mathcal{B}_{MN}(\mathbf{H}_l, \epsilon), & \text{if } l = 0. \end{cases}$$

The quantized random variable $\hat{\mathbf{S}}$ is defined as follows: $\hat{\mathbf{S}} = i$ if $\mathbf{S} \in \mathcal{S}_i$. The quantized random variables $\hat{\mathbf{U}}$, $\hat{\mathbf{Y}}$ and $\hat{H}$ are defined similarly. The joint probability of $\hat{\mathbf{S}}, \hat{\mathbf{U}}, \hat{\mathbf{Y}}, \hat{H}$ is,

$$P\left(\hat{\mathbf{S}} = i, \hat{\mathbf{U}} = j, \hat{\mathbf{Y}} = k, \hat{H} = l\right) = \int_{\mathbf{s}\in\mathcal{S}_i} \int_{\mathbf{u}\in\mathcal{U}_j} \int_{\mathbf{y}\in\mathcal{Y}_k} \int_{\mathbf{H}\in\mathcal{H}_l} f(\mathbf{s}) f(\mathbf{y}, \mathbf{H} \mid \mathbf{s}, \mathbf{x} = (\mathbf{s}, \mathbf{u})) q_G(\mathbf{u} \mid \mathbf{s}) \, d\mathbf{H} \, d\mathbf{y} \, d\mathbf{u} \, d\mathbf{s}$$

The Gelfand-Pinsker achievable rate corresponding to the quantized random variables is,

$$\hat{R} = \sum_{i,j,k,l} P\left(\hat{\mathbf{S}} = i, \hat{\mathbf{U}} = j, \hat{\mathbf{Y}} = k, \hat{H} = l\right) \log \frac{P(\hat{\mathbf{U}} = j \mid \hat{\mathbf{Y}} = k, \hat{H} = l)}{P(\hat{\mathbf{U}} = j \mid \hat{\mathbf{S}} = i)} \tag{48}$$

We claim that $\hat{R} = R_{\text{achievable}} + o_{\epsilon,\rho}(1)$ where $o_{\epsilon,\rho}(1)$ is a term that approaches 0 as $\epsilon \to 0$ and $\rho \to \infty$.

To see this, first note that when $\mathbf{s} \in \mathcal{S}_0$ or $\mathbf{u} \in \mathcal{U}_0$ or $\mathbf{y} \in \mathcal{Y}_0$ or $\mathbf{H} \in \mathcal{H}_0$, the contribution to $F(q_G, Q_G)$ in (11) approaches 0 as $\rho \to \infty$. In addition, $\log \frac{Q_G(\mathbf{u} \mid \mathbf{y}, \mathbf{H})}{q_G(\mathbf{u} \mid \mathbf{s})}$ is uniformly continuous in the region $\mathbf{s} \in \overline{\mathcal{S}_0}$, $\mathbf{u} \in \overline{\mathcal{U}_0}$, $\mathbf{y} \in \overline{\mathcal{Y}_0}$, $\mathbf{H} \in \overline{\mathcal{H}_0}$. Hence,

$$F(q_G, Q_G) = \sum_{i \neq 0, j \neq 0, k \neq 0, l \neq 0} P\left(\hat{\mathbf{S}} = i, \hat{\mathbf{U}} = j, \hat{\mathbf{Y}} = k, \hat{H} = l\right) \log \frac{Q_G(\mathbf{u}_j \mid \mathbf{y}_k, \mathbf{H}_l)}{q_G(\mathbf{u}_j \mid \mathbf{s}_i)} + o_{\epsilon,\rho}(1)$$

In addition, by the uniform continuity of the Gaussian distribution in the region $\mathbf{s} \in \overline{\mathcal{S}_0}$, $\mathbf{u} \in \overline{\mathcal{U}_0}$, $\mathbf{y} \in \overline{\mathcal{Y}_0}$, $\mathbf{H} \in \overline{\mathcal{H}_0}$,

$$\frac{P(\hat{\mathbf{U}} = j \mid \hat{\mathbf{Y}} = k, \hat{H} = l)}{Q_G(\mathbf{u}_j \mid \mathbf{y}_k, \mathbf{H}_l)} = 1 + o_{\epsilon,\rho}(1)$$

and

$$\frac{P(\hat{\mathbf{U}} = j \mid \hat{\mathbf{s}} = i)}{q_G(\mathbf{u}_j \mid \mathbf{s}_i)} = 1 + o_{\epsilon,\rho}(1)$$

Finally by arguments similar to those indicated above, the contribution of terms with $i = 0$ or $j = 0$ or $k = 0$ or $l = 0$ in (48) is negligible.

Hence we obtained the desired claim that $\hat{R} = R_{\text{achievable}} + o_{\epsilon,\rho}(1)$.

APPENDIX V
PROOF OF THEOREM 2

Our approach is the following. We begin with an assignment of variables for the LAFP achievable region. This means a set of variables $\mathbf{X}_1$,...,$\mathbf{X}_L$ that are not necessarily independent. A set of matrices $\mathbf{F}_1, ..., \mathbf{F}_L$ and a set of auxiliary random variables $\mathbf{U}_l = \mathbf{F}_l \mathbf{S}_l + \mathbf{X}_l$ where $\mathbf{S}_l = \Sigma_{i<l} \mathbf{X}_i$. Recall that in our current context, $\mathbf{X} = \mathbf{X}_1 + ... + \mathbf{X}_L$ denotes the transmitted symbol of the MIMO-BC channel, while $\mathbf{X}_l$ denotes the transmitted signal to user l, equivalent to $\mathbf{X}$ as in Sec. III-B.

We will construct an alternative set of independent random variables $\hat{\mathbf{X}}_1, ...\hat{\mathbf{X}}_L$ and $\hat{\mathbf{F}}_1, ..., \hat{\mathbf{F}}_L$ such that the transmitted signal $\hat{\mathbf{X}} \triangleq \hat{\mathbf{X}}_1 + ... + \hat{\mathbf{X}}_L = \mathbf{X}_1 + ... + \mathbf{X}_L = \mathbf{X}$. Thus, the distribution of the actual transmitted signal is unchanged and satisfies the power constraint. Furthermore, we show that for similarly defined $\hat{\mathbf{U}}_l = \hat{\mathbf{F}}_l \hat{\mathbf{S}}_l + \hat{\mathbf{X}}_l$ and $\hat{\mathbf{S}}_l = \Sigma_{i<l} \hat{\mathbf{X}}_i$, the achievable rates satisfy $\hat{R}_l \geq R_l$, where

$$\hat{R}_l \triangleq I(\hat{\mathbf{U}}_l; \mathbf{Y}_l, \mathbf{H}_l) - I(\hat{\mathbf{U}}_l; \hat{\mathbf{S}}_l), \qquad R_l = I(\mathbf{U}_l; \mathbf{Y}_l, \mathbf{H}_l) - I(\mathbf{U}_l; \mathbf{S}_l)$$

A. *Definition of* $\hat{\mathbf{X}}_1, ..., \hat{\mathbf{X}}_L$

For each $l = 1, ..., L$, using Gram-Schmidt orthogonalization, $\mathbf{X}_l$ can be written as $\mathbf{X}_l = \Gamma_l \mathbf{S}_l + \mathbf{X}'_l$ where Γ_l is a matrix and where $\mathbf{S}_l$ and $\mathbf{X}'_l$ are uncorrelated. Therefore, since we have assumed, in our definition of the LAFP region in Sec. IV-A, that all variables are jointly Gaussian, they are independent. With this

definition,

$$\begin{aligned}
\mathbf{X} &= \mathbf{S}_L + \mathbf{X}_L = (\mathbf{I} + \Gamma_L)\mathbf{S}_L + \mathbf{X}'_L \\
&= (\mathbf{I} + \Gamma_L)\left[\mathbf{S}_{L-1} + \mathbf{X}_{L-1}\right] + \mathbf{X}'_L = (\mathbf{I} + \Gamma_L)\left[(\mathbf{I} + \Gamma_{L-1})\mathbf{S}_{L-1} + \mathbf{X}'_{L-1}\right] + \mathbf{X}'_L \\
&\quad ... \\
&= (\mathbf{I} + \Gamma_L) \cdot ... \cdot (\mathbf{I} + \Gamma_2)\mathbf{X}'_1 + (\mathbf{I} + \Gamma_L) \cdot ... \cdot (\mathbf{I} + \Gamma_3)\mathbf{X}'_2 + ... + (\mathbf{I} + \Gamma_L)\mathbf{X}'_{L-1} + \mathbf{X}'_L
\end{aligned}$$

We thus define $\hat{\mathbf{X}}_l = \mathbf{G}_l\mathbf{X}'_l$ where $\mathbf{G}_l = (\mathbf{I} + \Gamma_L) \cdot ... \cdot (\mathbf{I} + \Gamma_{l+1})$ $l = 1, ..., L-1$, $\mathbf{G}_L = \mathbf{I}$. By construction, $\sum_{l=1}^{L} \hat{\mathbf{X}}_l = \mathbf{X} = \sum_{l=1}^{L} \mathbf{X}_l$, as desired.

The following lemma summarized some properties of our random variables.

Lemma 2: For all $l = 1, ..., L$,

1) $\hat{\mathbf{X}}_l$ is independent of $\mathbf{X}_1, ..., \mathbf{X}_{l-1}$.
2) $\hat{\mathbf{X}}_l$ is independent of $\mathbf{S}_1, ..., \mathbf{S}_l$.
3) $\hat{\mathbf{X}}_l$ is independent of $\hat{\mathbf{X}}_1, ..., \hat{\mathbf{X}}_{l-1}$.
4) $\hat{\mathbf{S}}_l = G_{l-1}\mathbf{S}_l$

Proof: To prove property 1, observe that the following Markov relations hold: $\mathbf{X}_1, ..., \mathbf{X}_{l-1} \longleftrightarrow \mathbf{S}_l \longleftrightarrow \mathbf{S}_l, \mathbf{X}_l \longleftrightarrow \hat{\mathbf{X}}_l$. $\hat{\mathbf{X}}_l$, by construction, is independent of $\mathbf{S}_l$. It is thus straightforward to verify, using this Markov relation, that it is also independent of $\mathbf{X}_1, ..., \mathbf{X}_{l-1}$. To obtain properties 2 and 3, observe that $\mathbf{S}_1, ..., \mathbf{S}_l$ and $\hat{\mathbf{X}}_1, ..., \hat{\mathbf{X}}_{l-1}$ are functions of $\mathbf{X}_1, ..., \mathbf{X}_{l-1}$ and thus are independent of $\hat{\mathbf{X}}_l$.

The last property is easily obtained by induction. For $l = 1$,

$$\hat{\mathbf{S}}_1 = \sum_{i<1} \hat{\mathbf{X}}_i = \mathbf{0} = \sum_{i<1} \mathbf{X}_i = \mathbf{S}_1$$

The rest is obtained by the following induction:

$$\begin{aligned}
\hat{\mathbf{S}}_{l+1} &= \hat{\mathbf{S}}_l + \hat{\mathbf{X}}_l = G_{l-1}\mathbf{S}_l + G_l\mathbf{X}'_l = G_l(I + \Gamma_l)\mathbf{S}_l + G_l\mathbf{X}'_l = G_l\left[(I + \Gamma_l)\mathbf{S}_l + \mathbf{X}'_l\right] \\
&= G_l\left[\mathbf{S}_l + \Gamma_l\mathbf{S}_l + \mathbf{X}'_l\right] = G_l\left[\mathbf{S}_l + \mathbf{X}_l\right] = G_l\mathbf{S}_{l+1}
\end{aligned}$$

□

B. Definition of $\hat{\mathbf{F}}_1, ..., \hat{\mathbf{F}}_L$

We have not yet defined $\hat{\mathbf{F}}_l$. To do so, we first consider $\mathbf{G}_l \cdot \mathbf{U}_l$. By the definition of $\mathbf{U}_l$

$$\mathbf{G}_l \cdot \mathbf{U}_l = \mathbf{G}_l[\mathbf{F}_l\mathbf{S}_l + \mathbf{X}_l] = \mathbf{G}_l[(\mathbf{F}_l + \Gamma_l)\mathbf{S}_l + \mathbf{X}'_l] = \mathbf{G}_l(\mathbf{F}_l + \Gamma_l)\mathbf{S}_l + \hat{\mathbf{X}}_l \tag{49}$$

where the last inequality was obtained by the definition of $\hat{\mathbf{X}}_l$, above. Using Gram-Schmidt orthogonalization, $\mathbf{S}_l$ can be written as $\mathbf{S}_l = \mathbf{B}_l \cdot \hat{\mathbf{S}}_l + \mathbf{D}_l$ where $\mathbf{B}_l$ is a matrix and $\mathbf{D}_l$ is uncorrelated with $\hat{\mathbf{S}}_l$. Since the variables are jointly Gaussian, $\mathbf{D}_l$ is also independent of $\hat{\mathbf{S}}_l$. We proceed

$$\begin{aligned}
\mathbf{G}_l \cdot \mathbf{U}_l &= \mathbf{G}_l(\mathbf{F}_l + \Gamma_l)[\mathbf{B}_l\hat{\mathbf{S}}_l + \mathbf{D}_l] + \hat{\mathbf{X}}_l \\
&= \mathbf{G}_l(\mathbf{F}_l + \Gamma_l)\mathbf{B}_l\hat{\mathbf{S}}_l + \hat{\mathbf{X}}_l + \mathbf{G}_l(\mathbf{F}_l + \Gamma_l)\mathbf{D}_l
\end{aligned} \tag{50}$$

We define $\hat{\mathbf{F}}_l = \mathbf{G}_l(\mathbf{F}_l + \Gamma_l)\mathbf{B}_l$.

C. Proof of $\hat{R}_l > R_l$

Recall that $\hat{\mathbf{U}}_l = \hat{\mathbf{F}}_l\hat{\mathbf{S}}_l + \hat{\mathbf{X}}_l$. To prove $\hat{R}_l > R_l$, we first define an intermediate auxiliary variable $\tilde{\mathbf{U}}_l = \mathbf{G}_l \cdot \mathbf{U}_l$. Since $\tilde{\mathbf{U}}_l$ is a function of $\mathbf{U}_l$, we have

$$\begin{aligned}
R_l &= I(\tilde{\mathbf{U}}_l, \mathbf{U}_l; \mathbf{Y}_l, H_l) - I(\tilde{\mathbf{U}}_l, \mathbf{U}_l; \mathbf{S}_l) \\
&= H(\tilde{\mathbf{U}}_l, \mathbf{U}_l \mid \mathbf{S}_l) - H(\tilde{\mathbf{U}}_l, \mathbf{U}_l \mid \mathbf{Y}_l, H_l) \\
&= H(\tilde{\mathbf{U}}_l \mid \mathbf{S}_l) - H(\tilde{\mathbf{U}}_l \mid \mathbf{Y}_l, H_l) + H(\mathbf{U}_l \mid \tilde{\mathbf{U}}_l, \mathbf{S}_l) - H(\mathbf{U}_l \mid \tilde{\mathbf{U}}_l, \mathbf{Y}_l, H_l) \\
&= \Big[H(\tilde{\mathbf{U}}_l \mid \mathbf{S}_l) - H(\tilde{\mathbf{U}}_l \mid \mathbf{Y}_l, H_l)\Big] + \Big[I(\mathbf{U}_l; \tilde{\mathbf{U}}_l, \mathbf{Y}_l, H_l) - I(\mathbf{U}_l; \tilde{\mathbf{U}}_l, \mathbf{S}_l)\Big]
\end{aligned}$$

We now wish to show that the contents of the second brackets are non-positive. For this purpose, we will show that the following Markov relations hold: $\mathbf{U}_l \longleftrightarrow \tilde{\mathbf{U}}_l, \mathbf{S}_l \longleftrightarrow \tilde{\mathbf{U}}_l, \hat{\mathbf{X}}_l, \hat{\mathbf{S}}_l \longleftrightarrow \tilde{\mathbf{U}}_l, \mathbf{X} \longleftrightarrow \tilde{\mathbf{U}}_l, \mathbf{Y}_l, H_l$. The desired result will then follow from the first and last Markov relations, using the data processing inequality.

The second relation (first Markov triple) follows from the fact that $\hat{\mathbf{X}}_l$ and $\hat{\mathbf{S}}_l$ may be determined from $\tilde{\mathbf{U}}_l$ and $\mathbf{S}_l$ by means of deterministic functions: $\hat{\mathbf{X}}_l$ through (49), and $\hat{\mathbf{S}}_l$, by Lemma 2 satisfies $\hat{\mathbf{S}}_l = \mathbf{G}_{l-1}\mathbf{S}_l$. For the third relation, observe that $\mathbf{X} = \hat{\mathbf{S}}_l + \hat{\mathbf{X}}_l + \sum_{i>l} \hat{\mathbf{X}}_i$. By the above definition all $\{\hat{\mathbf{X}}_i\}_{i>l}$, are independent of $\mathbf{U}_l, \tilde{\mathbf{U}}_l, \hat{\mathbf{S}}_l, \mathbf{S}_l$ and $\hat{\mathbf{X}}_l$. Therefore this Markov relation holds. The last Markov relation is straightforward.

We thus have,

$$R_l \leq H(\tilde{\mathbf{U}}_l \mid \mathbf{S}_l) - H(\tilde{\mathbf{U}}_l \mid \mathbf{Y}_l, H_l) \tag{51}$$

Examining the first element of the above difference, we obtain:

$$\begin{aligned}
H(\tilde{\mathbf{U}}_l \mid \mathbf{S}_l) &= H(\mathbf{G}_l(\mathbf{F}_l + \Gamma_l)\mathbf{S}_l + \hat{\mathbf{X}}_l \mid \mathbf{S}_l) = H(\hat{\mathbf{X}}_l \mid \mathbf{S}_l) = H(\hat{\mathbf{X}}_l) = H(\hat{\mathbf{X}}_l \mid \hat{\mathbf{S}}_l) = H(\hat{\mathbf{F}}_l\hat{\mathbf{S}}_l + \hat{\mathbf{X}}_l \mid \hat{\mathbf{S}}_l) \\
&= H(\hat{\mathbf{U}}_l \mid \hat{\mathbf{S}}_l) \qquad (52)
\end{aligned}$$

where the first equality follows from the definition of $\tilde{\mathbf{U}}_l$ and from (49). The third equality follows from the independence of $\hat{\mathbf{X}}_l$ and $\mathbf{S}_l$ and the fourth from the independence of $\hat{\mathbf{X}}_l$ and $\hat{\mathbf{S}}_l$.

Examining the second element of (51), we have

$$\begin{aligned}
H(\tilde{\mathbf{U}}_l \mid \mathbf{Y}_l, H_l) &= H(\hat{\mathbf{U}}_l + \mathbf{G}_l(\mathbf{F}_l + \Gamma_l)\mathbf{D}_l \mid \mathbf{Y}_l, H_l) \geq H(\hat{\mathbf{U}}_l + \mathbf{G}_l(\mathbf{F}_l + \Gamma_l)\mathbf{D}_l \mid \mathbf{Y}_l, H_l, \mathbf{D}_l) \\
&= H(\hat{\mathbf{U}}_l \mid \mathbf{Y}_l, H_l, \mathbf{D}_l) = H(\hat{\mathbf{U}}_l \mid \mathbf{Y}_l, H_l) \qquad (53)
\end{aligned}$$

The first equality follows from (50) and the definitions of $\tilde{\mathbf{U}}_l$, $\hat{\mathbf{U}}_l$ and $\hat{\mathbf{F}}_l$. The inequality results from the fact that conditioning cannot increase the entropy. To prove the last equality, we wish to show that $\mathbf{D}_l$ and $\hat{\mathbf{U}}_l$ are independent, given $\mathbf{Y}_l$ and H_l.

$\hat{\mathbf{U}}_l$ is a function of $\hat{\mathbf{S}}_l$ and $\hat{\mathbf{X}}_l$. Therefore, it suffices to show that $\mathbf{D}_l$ is independent of these two random variables, given $\mathbf{Y}_l$ and H_l. $\mathbf{D}_l$ is independent of $\hat{\mathbf{S}}_l$ by construction. In addition, $\hat{\mathbf{X}}_l$ is independent of

$\mathbf{S}_l, \hat{\mathbf{S}}_l$ and $\mathbf{D}_l$, because $\mathbf{D}_l$ is a function of $\mathbf{S}_l$ and of $\hat{\mathbf{S}}_l$, where $\hat{\mathbf{S}}_l = \mathbf{G}_{l-1}\mathbf{S}_l$ (by Lemma 2), and $\hat{\mathbf{X}}_l$ is independent of $\mathbf{S}_l$ (again, by Lemma 2). Therefore, $\mathbf{D}_l$ is independent of $\hat{\mathbf{S}}_l$ and $\hat{\mathbf{X}}_l$. To show that the independence is maintained even when we condition by $\mathbf{Y}_l$ and H_l, we prove the following Markov chain relation $\mathbf{D}_l \longleftrightarrow \hat{\mathbf{S}}_l, \hat{\mathbf{X}}_l \longleftrightarrow \hat{\mathbf{S}}_l, \hat{\mathbf{X}}_l, ...\hat{\mathbf{X}}_L \longleftrightarrow \mathbf{X} \longleftrightarrow H_l, \mathbf{Y}_l$. The second relation (first Markov triple) holds because the random variables $\hat{\mathbf{X}}_{l+1}, ..., \hat{\mathbf{X}}_L$ are independent of $\hat{\mathbf{S}}_l$ and of $\hat{\mathbf{X}}_l$ by Lemma 2, and of $\mathbf{D}_l$, by virtue of it being a function of $\hat{\mathbf{S}}_l$ and $\mathbf{S}_l$. The third relation holds because $\mathbf{X} = \hat{\mathbf{S}}_l + \hat{\mathbf{X}}_l + ... + \hat{\mathbf{X}}_L$. The fourth relation holds because $\mathbf{Y}_l = H_l\mathbf{X} + \mathbf{Z}_l$ and H_l and $\mathbf{Z}_l$ are independent of the other random variables $\mathbf{D}_l, \hat{\mathbf{S}}_l, \hat{\mathbf{X}}_l, ...\hat{\mathbf{X}}_L, \mathbf{X}$.

Combining (51), (52) and (53) we obtain,

$$R_l \leq H(\hat{\mathbf{U}}_l \mid \hat{\mathbf{S}}_l) - H(\hat{\mathbf{U}}_l \mid \mathbf{Y}_l, H_l) = I(\hat{\mathbf{U}}_l; \mathbf{Y}_l, H_l) - I(\hat{\mathbf{U}}_l; \hat{\mathbf{S}}_l) = \hat{R}_l$$

This completes the proof. □

Appendix VI

Computing the LAFP Achievable Region when the Number of Users is Greater than Two

In Sec. IV-B we considered the computation of the LAFP achievable region over a fading MIMO-BC channel where the number of users is two. In this appendix we briefly consider the case of more than two users. To obtain the LAFP achievable region, we could again (as in Sec. IV-B) apply a grid search to obtain $\{\Sigma_X^{(l)}\}_{l=1}^L$. A straightforward approach would be to compute, for each choice of such matrices, the achievable rates for each of the individual users by selecting the matrices $\mathbf{F}$, for each user (except for the first who does not have an associated $\mathbf{F}$ matrix) so as to maximize (35). However, the computational complexity of such an approach would grow exponentially with the number of users.

The following observation can be used to reduce the number of computations. The achievable rate for user l is a function of $\Sigma_X^{(l)}$ (the covariance matrix of its transmitted signal $\mathbf{X}_l$), of $\Sigma_S^{(l)} \triangleq \sum_{i<l} \Sigma_X^{(i)}$ (the covariance matrix of the interference $\mathbf{S}_l = \sum_{i<l} \mathbf{X}_i$) and $\Sigma_Z^{(l)} \triangleq \sum_{i>l} H \Sigma_X^{(i)} H^T + \mathbf{I}$ (the covariance matrix of the effective noise $\mathbf{Z}_l = H \sum_{i>l} \mathbf{X}_i + \mathbf{Z}$). Thus, the achievable rate for user l needs to be computed only once for each of the possible choices of $\Sigma_S^{(l)}$, $\Sigma_X^{(l)}$ and $\Sigma_Z^{(l)}$, and not for each choice of $\{\Sigma_X^{(i)}\}_{i=1}^L$. A dynamic-programming algorithm that relies on this observation can dramatically reduce the number of computations. This approach is useful when the number of transmit antennas and the number of receive antennas of each user is small (the number of users can be large). Otherwise we can resort to suboptimal methods for computing the transmit covariances $\{\Sigma_X^{(l)}\}_{l=1}^L$ (and the $\mathbf{F}$ matrices), e.g. using gradient descent or alternate maximization that maximizes the sum rate with respect to two $\Sigma_X^{(l)}$-s at a time, while fixing the other $\Sigma_X^{(l)}$-s.

Appendix VII

Computing a Cooperative Upper-Bound in our Setting

Sato's upper bound [23] on the sum rate capacity (the maximum achievable sum-rate) of a broadcast channel relies on two observations:

1) A fundamental assumption in the broadcast channel model is that the users are not able to cooperate in their decoding. Consider a virtual channel where the users are allowed to cooperate. The sum capacity in this channel is clearly an upper bound on the sum rate capacity of the true channel. Such a cooperative model is equivalent to transmission to a single virtual user, to whom all the outputs of the broadcast channel users are made available.
2) The capacity region of a broadcast channel depends not on the joint distribution $\Pr(\mathbf{Y}_1, H_1, ..., \mathbf{Y}_L, H_L \mid \mathbf{X})$ but on the marginal distributions $\Pr(\mathbf{Y}_1, H_1 \mid \mathbf{X}), ..., \Pr(\mathbf{Y}_L, H_L \mid \mathbf{X})$ alone. Thus, we may alter our model by introducing correlation between the noise signals and channel matrices of different users. As long as the marginal statistics of the individual channels to each of the users stay the same, the resulting broadcast channel's capacity region will remain unchanged. However, introducing correlations *could* alter (and tighten) the above-mentioned cooperative upper bound.

Note that with any valid choice of correlation that we choose to introduce, the maximum cooperative sum-rate produces an upper bound on the broadcast channel's sum-rate capacity. We refer to such an upper bound as a *cooperative* upper bound. The Sato upper bound is the tightest such bound.

Consider the channel to the virtual single user corresponding to the fading MIMO-BC example of Sec. IV-B. This user will observe a virtual channel matrix and a virtual noise defined as,

$$H = \begin{bmatrix} H^{(1)} \\ H^{(2)} \end{bmatrix} \quad \text{and} \quad \mathbf{Z} = \begin{bmatrix} Z^{(1)} \\ Z^{(2)} \end{bmatrix}$$

Our above discussion implies that we may freely introduce correlations as long as we do not alter the statistics of the channel observed by each of the individual users. We may thus introduce a correlation between the two noise signals $Z^{(1)}$ and $Z^{(2)}$, following the examples of [6] and [29]. We may also introduce correlation between the two channel matrices $H^{(1)}$ and $H^{(2)}$. Furthermore, we may introduce correlation between the channel matrix of one user and the noise of the other.

The possible values for H are,

$$H \in \left\{ \mathbf{H}_1 = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 1 \end{bmatrix}, \mathbf{H}_2 = \begin{bmatrix} 1 & 0.4 \\ 3 & 1 \end{bmatrix}, \mathbf{H}_3 = \begin{bmatrix} 1 & 3 \\ 0.4 & 1 \end{bmatrix}, \mathbf{H}_4 = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \right\}$$

Let $p(H)$ denote the probability assignment to each of the above matrices. To preserve the marginal statistics of the channel to each of the individual users, we require that $p(H)$ satisfy the following constraints,

$$p(\mathbf{H}_1) + p(\mathbf{H}_2) = \frac{1}{2}$$
$$p(\mathbf{H}_1) + p(\mathbf{H}_3) = \frac{1}{2}$$

Furthermore, for $p(H)$ to be a valid probability assignment, it must satisfy,

$$p(\mathbf{H}_1) + p(\mathbf{H}_2) + p(\mathbf{H}_3) + p(\mathbf{H}_4) = 1$$

The constraints imply that $p(H)$ is completely described by $\alpha \triangleq p(\mathbf{H}_1)$. That is, for any $\alpha \in [0, 1/2]$, we have

$$p(\mathbf{H}_1) = p(\mathbf{H}_4) = \alpha, \quad p(\mathbf{H}_2) = p(\mathbf{H}_3) = \frac{1}{2} - \alpha$$

One way to introduce correlation between the various noise elements is to follow the approach of [6]. That is, introduce a correlation coefficient $\rho \in (-1, 1)$ and consider a virtual noise $\mathbf{Z}$ whose covariance matrix is,

$$\mathrm{Cov}(\mathbf{Z}) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

However, a more general approach would introduce correlation between the virtual noise and the above virtual channel matrix in the following way: We will consider *four* correlation coefficients $\rho_1, \rho_2, \rho_3, \rho_4$ such that,

$$\mathrm{Cov}(\mathbf{Z} \mid H = \mathbf{H}_i) = \begin{bmatrix} 1 & \rho_i \\ \rho_i & 1 \end{bmatrix} \quad i = 1, ..., 4$$

The channel noise observed by each of the users remains distributed as $\mathcal{N}(0, 1)$. Furthermore, each of the individual realizations of $Z^{(1)}$ and $Z^{(2)}$ remains independent of the respective channel matrices $H^{(1)}$ and $H^{(2)}$. Thus, the marginal statistics of the channels to each of the individual users remain unchanged, as desired.

The capacity of the channel to the virtual user is now obtain by taking the maximum of,

$$I(\mathbf{X}; \mathbf{Y}, H) = I(\mathbf{X}; H) + I(\mathbf{X}; \mathbf{Y} \mid H) = I(\mathbf{X}; \mathbf{Y} \mid H) = \sum_{i=1}^{4} p(\mathbf{H}_i) I(\mathbf{X}; \mathbf{Y} \mid H = \mathbf{H}_i)$$

The first equality is obtained by the chain rule for mutual information, and the second by the independence of $\mathbf{X}$ and H. The distribution that maximizes the above is clearly Gaussian. Thus,

$$C = \max_{\Sigma_X} \sum_{i=1}^{4} p(\mathbf{H}_i) \frac{1}{2} \log\det\left(\mathbf{I} + \Lambda_i^{-1/2} \mathbf{H}_i \Sigma_X \mathbf{H}_i^T \Lambda_i^{-1/2}\right) \tag{54}$$

where $\Lambda_i \triangleq \mathrm{Cov}(\mathbf{Z} \mid H = \mathbf{H}_i)$.

We may now numerically obtain a cooperative upper bound in the following way. We consider all choices of $\alpha, \rho_1, ..., \rho_4$ along a fine grid. For each such choice, we evaluate (54) by applying semidefinite programming to determine the Σ_X that achieves the maximum. Each choice of $\alpha, \rho_1, ..., \rho_4$ produces a cooperative bound. We conclude by selecting the lowest (tightest) bound[11].

In our numerical results (as presented in Sec. IV-B), the tightest bound was obtained by setting $\alpha = 0$ and $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0.3$. Thus, the tightest bound was obtained with a limited exploitation of the available degrees of freedom in the above approach.

ACKNOWLEDGEMENTS

We would like to thank Marc Teboulle for helpful discussions, and the anonymous reviewers for their comments, that helped improve the presentation of the paper.

[11]Note that the bound obtained in this way is not necessarily the true Sato upper bound (i.e., the tightest possible cooperative bound), because we have not proven that our approach exhausts all the possible ways of introducing valid correlations between the various signals.

REFERENCES

[1] A. Bennatan, D. Burshtein, G. Caire and S. Shamai "Superposition Coding for Side-Information Channels," *IEEE Trans. Inform. Theory*, vol 52, pp. 1872–1889, May 2005.

[2] P. Billingsley, *Probability and measure*, John Wiley and Sons, 3rd ed. 1995.

[3] S. Borade and L. Zheng, "Writing on fading paper and causal transmitter CSI," available online in http://arxiv.org/abs/cs.IT/0511081.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[5] G. Caire, Talk at Technion, Israel, Jan. 2006.

[6] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform, Theory*, Vol. 49, No. 7, pp. 1691–1706, July 2003.

[7] A. S. Cohen and A. Lapidoth, "The Gaussian Watermarking Game," *IEEE Trans. on Inform. Theory* vol. 48, no. 6, June 2002.

[8] M. Costa, "Writing on dirty paper," *IEEE Trans. on Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[9] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

[10] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT–18, no. 1, pp. 2–14, Jan. 1972.

[11] F. Dupuis, W. Yu and F. M.. J. Willems "Blahut-Arimoto Algorithms for Computing Channel Capacity and Rate-Distortion With Side Information," *IEEE International Symposium on Information Theory (ISIT)*, Chicago, USA. 2004.

[12] A. El Gamal,"The Capacity of a Class of Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. IT–25, no. 2, pp. 166–169, Mar. 1979.

[13] U. Erez, S. Shamai and R. Zamir, "Capacity and Lattice-Strategies for Cancelling Known Interference," *IEEE Trans. on Inform. Theory*, November 2005.

[14] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, January 1980.

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1988.

[16] S. A. Jafar and A. J. Goldsmith, "Isotropic Fading Vector Broadcast Channels: The Scalar Upper Bound and Loss in Degrees of Freedom," *IEEE Trans. on Inform. Theory*, vol. 51, No. 3 pp. 848–857 Mar. 2005.

[17] C.Heegard and A. El Gamal, "On the Capacity of Computer Memory with Defects," *IEEE Trans. Inform. Theory*, Vol. 29, No. 5, Sep. 1983.

[18] N. Jindal, "MIMO Broadcast Channels with Finite Rate Feedback", Submitted to *IEEE Trans. on Inform. Theory*.

[19] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*, Englewood Cliffs, NJ : PTR Prentice-Hall, 1993.

[20] A. Kusnetsov and B. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered. Inform.*, vol. 10, no. 2, pp. 52–60, 1974.

[21] A. Lapidoth, S. Shamai (Shitz) and M. A. Wigger, "On the Capacity of Fading MIMO Broadcast Channels with Imperfect Transmitter Side-Information," *Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing*, Sept. 28-30, 2005.

[22] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd Edition, McGraw-Hill.

[23] H. Sato, "An outer bound on the capacity region of broadcast channel," *IEEE. Trans. Inform. Theory*, vol. IT–24, pp. 374-377, May 1978.

[24] C. Shannon, "Channels with side information at the transmitter," *IBM J. Res. & Dev.*, pp. 289–293, 1958.

[25] M. Sharif and B. Hassibi "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. on Inform. Theory*, vol. IT–51, pp. 506–522, Feb. 2005.

[26] R. A. Tapia and M. W. Trosset, "An extension of the Karush-Kuhn-Tucker necessity conditions to infinite programming," SIAM Review Vol. 36, No. 1, pp. 1–17, March 1994.

[27] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585-596, Nov. 1999.

[28] D. Tuninetti and S. Shamai, “Fading Gaussian broadcast channels with state information at the receivers,” *Advances in Network Information Theory, The DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. Piscataway, NJ: Rutgers Univ. Press, 2003, vol. 66, pp. 139-150.

[29] S. Vishwanath, N. Jindal, A. Goldsmith, “Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels,” *IEEE Trans. on Inform. Theory*, vol. 49, pp. 2658–2668 Oct. 2003.

[30] H. Weingarten, Y. Steinberg and S. Shamai (Shitz), “The Capacity Region of the Gaussian MIMO Broadcast Channel,” *38th Annual Conference on Information Sciences and System* (CISS 2004), March 14–19, 2004, Princeton, N.J., USA.

[31] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, Writing on colored paper, in Proc. IEEE Int. Symp. Information Theory, Washington, DC, June 2001, p. 302.

[32] W. Yu and J. M. Cioffi, ”Sum Capacity of Gaussian Vector Broadcast Channels,” *IEEE Trans. on Inform. Theory*, Vol. 50, No. 9, pp. 1875–1892, Sep. 2004 .

[33] R. Zamir, S. Shamai, and U. Erez, “Nested linear/lattice codes for structured multiterminal binning,” *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

LIST OF TABLES

(None)

LIST OF FIGURES